Author:
Hodges, Christopher John

Title:
An optical study of III-nitride semiconductor devices, their thermal properties and degradation mechanisms

An optical study of III-nitride semiconductor devices, their thermal properties and degradation mechanisms

Author:
Christopher John Hodges

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of
Doctor of Philosophy
in the
Faculty of Science

June 2014

Word count: 51 015

Abstract

Degradation processes in AlGaN/GaN high electron mobility transistors were investigated by optical methods. Temperatures within device channels, as well as electric fields are important for on-state degradation. Raman and photoluminescence (PL) thermography were used to investigate these temperatures and the thermal conductivity of GaN channels in both conventional GaN-on-SiC structures and novel AlGaN/GaN/AlGaN-on-Si double heterostructure field effect transistors (DHFETs). The thin (150 nm) GaN channel layer in the DHFET had a lower thermal conductivity, at $\sim 60\ \mathrm{W\,m^{-1}\,K^{-1}}$ than typical epilayers, which at ~ 2 µm thick have more than twice this value. This reduced thermal conductivity has implications for the design of devices employing thin GaN layers, especially when combined with the thick strain relief layers common on Si substrates, as the resulting high temperatures will affect their reliability by on- state thermal degradation processes. The depth resolution of Raman thermography on devices with typical GaN buffers usually limits results to one temperature, averaged through the buffer thickness. A method was developed to improve the depth resolution using a spatial filter and azimuthal polarisation; when combined with offset focal planes it was possible to obtain temperatures of the top and bottom of the GaN epilayer separately.

Off-state degradation processes are more closely related to electric fields than self-heating; the generation of leakage current paths from the gate to the channel is particularly important. This leakage-path generation and associated localised electroluminescence (EL) emission was studied using EL imaging and spectroscopy combined with deep UV PL spectroscopy. The PL from the AlGaN barrier was reduced in regions associated with localised EL, indicating the formation of non-radiative recombination centres in the form of defects in the AlGaN. These non-radiative recombination centres were found to be generated over a larger area than the location of the gate leakage currents – these currents only start to flow when sufficient defects form to constitute a path.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed: .....

Date: .....

Acknowledgements

While my name may appear on the front of this thesis, there are many people without whom I could not have even come close to writing it. First I would like to thank Martin Kuball, my supervisor, who not only gave me the opportunity to carry out this work, but supported me throughout it. The people, past and present, of the Centre for Device Thermography and Reliability, and the wider Microstructures group have been excellent colleagues providing not just stimulating discussions on technical and many other matters, but also cake in many forms and a most supportive working environment. On scientific matters particular thanks are due to James Pomeroy, Miguel Montes Bajo, Julián Anaya Calvo, Nicole Killat, Huarui Sun, Mike Uren and Janina Möreke – the discussions I have had with them have cleared up many questions. The whole group have been thoroughly supportive and willing to share their considerable knowledge, I could easily list all their names. The Physics workshop have been of great help turning my sketches into functioning reality in time for me to get results.

Outside the University I would like to thank Daniel Wolverson at the University of Bath not just for the use of his spectrometer but for his input into the experiments for which I used it. I would like to thank the many collaborators from IMEC, MIT and UCSB who supplied me with samples in particular Denis Marcon and Steve Stoffels of IMEC for their patience as I used and abused their devices for purposes they hadn't envisaged. Some of my former colleagues at Renishaw, particularly Dave Cutler and Bob Bennett, are partially responsible for convincing me that I should embark on further academic study, for which I'd like to take this chance to thank them. Of course, nothing can be done without funding, for which I would like to thank the EPSRC. Finally the support of my whole family has been absolutely essential and I would particularly like to thank them for their help and encouragement while I've been writing this thesis.

Contents

List of Figures

List of Tables

List of abbreviations

2DEG	two-dimensional electron gas
AFM	atomic force microscopy (or —microscope)
CCD	charge-coupled device
CMOS	complementary metal-oxide semiconductor
DC	direct current
DHFET	double heterostructure field effect transistor
DUT	device under test
DUV	deep ultraviolet
EL	electroluminescence
FET	field effect transistor
FIB	focussed ion beam
HEMT	high electron mobility transistor
HFET	heterostructure field effect transistor (or hetero*junction*—)
LED	light emitting diode
MFP	mean free path
PMT	photomultiplier tube
RMS	root-mean-square
$\mathbf{R_{on}}$	on-resistance
TEM	transmission electron microscopy (or —microscope)

Chapter 1

Introduction

The drive for more compact, powerful and efficient electronic systems, whether for communications, remote sensing (for example radar), or power supply systems has led to many advances in electronic devices, including some of the first applications of semiconductors. Radio communications and remote sensing operate (by definition) at radio frequency (RF), which is broadly defined as ranging from kHz to GHz. Most modern radio applications tend towards the higher end of this range, from MHz up to the THz boundary. Power supply switching frequencies tend to be lower, no more than a few MHz; while they still fall within the RF frequency range a distinction between applications is also implied in the terminology. The earliest descriptions of the transistor, from 1948, discuss its use in RF amplification.[1] Over the decades since then, all transistor operating parameters have improved markedly, driven by improvements in materials and manufacturing techniques, as well as new device designs.[2] Despite this progress, or perhaps even because of it, the demand for ever better transistor performance has not reduced. This is particularly true at high frequencies – GHz and above – but even at the lower frequencies used in switching power conversion, a transistor which can control more current, at a higher voltage, and with lower losses, can lead to tangible benefits for the system designer and user. Depending on the application, the current leading materials systems for RF amplification or switching power conversion may be gallium arsenide, silicon or silicon carbide, all of which leave room for an improved technology.

The compound semiconductors gallium nitride (GaN), aluminium nitride (AlN), and indium nitride (InN), along with the ternary compounds AlGaN and InGaN are collectively known as the III-nitrides. Their bandgaps range from 6.2 eV for AlN to 0.7 eV for InN. This makes the III-nitrides versatile for opto-electronic systems, as first demonstrated by Nakamura in 1991,[3] where they have proved their worth in applications as diverse as blue laser diodes and white light emitting diodes (LEDs), both of which are

now ubiquitous in consumer applications. They are also of interest for solar cells, where their direct bandgaps are favourable. In the visible and near infrared spectral ranges InN and InGaN are used. GaN, AlGaN, and AlN are known as wide-bandgap materials, with bandgaps at ultraviolet wavelengths – it is structures based on these that are of relevance to the work described in this thesis. While the optical applications of III-nitride electronics devices are widely known, here it is their use in transistors that is of interest, specifically in the form of high electron mobility transistors (HEMTs) – field effect transistors which enhance their operating parameters compared to those of bulk materials by confining current to a two-dimensional channel at the interface of two semiconductors with different bandgaps. These were originally developed in 1979,[4] using GaAs and AlGaAs; only in 1993 was a GaN-based HEMT developed[5] although GaN crystals had already been grown and their electrical properties studied in 1969.[6] The wide bandgaps of GaN and AlGaN enable high voltage and high temperature operation. With respect to more established transistor designs, the confinement of the electrons in the channel of a HEMT, by maximising the electron mobility, reduces the on-resistance (R_{on}) and therefore the power lost when the channel is conducting. Fast switching times between on- and off-states reduce the power losses due to switching. Reducing the power loss in the device itself and allowing high current densities at high voltages mean that GaN based transistors aim to address the requirements for more power, in a smaller space, with less requirement for cooling, in applications as diverse as electric cars, power conversion and radar. As they are radiation-hard compared to competing technologies, GaN HEMTs are of particular interest for space-based applications, such as satellite communications. The design of GaN-based HEMTs may currently be optimised to give 40 $W\,mm^{-1}$ of output power,[7] 1.9 kV breakdown voltage,[8] or a maximum oscillator frequency of 400 GHz for experimental devices. Commercial specifications are rather more cautious, but devices are available for RF amplification which can provide 100 W at 3.5 GHz and 50 V.[9] Hybrid power-switching devices using AlGaN/GaN HEMTs combined with Si field effect transistors (FETs) are sold to switch 600 V at 17 A.[10] AlGaN/GaN HEMTs therefore look set to continue the successes of III-nitride technology into new fields.

Although GaN-based transistors have reached the market, before widespread adoption there are still hurdles to be overcome. Two major challenges are cost and reliability, which are interlinked – the most reliable technology is rarely the most affordable, or indeed the most efficient or compact. In fact the cost-benefit position of AlGaN/GaN electronic devices depends on being able to achieve the maximum performance that the technology

offers, for a lifetime of years or decades. The increase in lifetime over the twenty years since the first AlGaN/GaN transistors has been remarkable – devices reported in 2009 last 10^5 times longer than early prototypes: many decades under operating conditions instead of a few hours.[11] This is in addition to steady increases in the maximum device current, operating voltage and usable channel temperature. However to obtain the maximum lifetime often means derating the operating parameters. Of particular relevance to these challenges for GaN-based devices is the typical need for manufacture by heteroepitaxy on a substrate of a different material, as growing GaN thick enough to be self-supporting is problematic. When making such bulk GaN wafers a substrate is still required for the GaN growth, but the substrate is removed before device fabrication. The selection of substrate for device growth is critical for both cost and reliability – GaN transistors would possibly be widespread already if the most cost-effective devices grown on Si could be made as good as the best devices grown on SiC. Even if this were achieved, the demand for improved performance would not stop. A third challenge relevant to the adoption of AlGaN/GaN power-switching HEMTs is that HEMTs are still commonly normally-on devices, which are unsuitable for some power supply designs. The Si FET in the GaN-Si hybrid construction (cascode[12] topography) mentioned above is used to overcome this, with some loss of efficiency. The ongoing development of a true normally-off HEMT on a Si substrate led to the device studied in Chapter 4.

To make further progress in improving this technology – which has already proved its worth – requires a deeper understanding of the mechanisms which limit device performance and especially device lifetime. There are two main drivers of degradation in AlGaN/GaN HEMTs: thermal ageing and electric field effects. They are not independent, but it is possible to consider some degradation mechanisms in terms of predominantly one or the other. If the thermal properties of a device are less than optimal, high temperatures in operation cause damage, prematurely ageing the device. To reduce this effect on device lifetime, the maximum current density may be restricted to keep the peak temperature within limits, reducing some of the benefits of using a GaN-based device in the first place. The effect of electric-field-related failure mechanisms is to limit the maximum voltage that the device can withstand if it is to survive for a reasonable lifetime. Once again, this reduces the benefits of GaN over competing technologies.

This thesis sets out to investigate a selection of these degradation mechanisms, and the thermal properties that play a part in them, in a range of devices grown on a variety

of substrates, and aimed at a variety of applications. The optical methods developed in this work, along with many similar techniques, are particularly well-suited to work on wide-bandgap semiconductors, which are transparent to visible light. By working with visible light, it is possible to monitor a device as it operates, on length scales comparable to the wavelength of the light used, while having no effect on the device under test. These length scales are highly compatible with the geometry of typical AlGaN/GaN HEMTs, in which features range from less than one micron to several microns in length, and up to hundreds of microns in width. With Raman thermography, for example, it is possible to study submicron variations in temperature, as a device is operated under harsh conditions. Alternatively, light may be emitted from a device in operation – electroluminescence – the spatial distribution and spectrum of which can give an insight into electric fields and current paths within the device. Specifically for this work, electroluminescence and photoluminescence spectroscopy combined with electroluminescence imaging and electrical logging are used to study off-state stress, which is field driven due to the lack of current for self heating. The thermal properties of GaN epilayers and their connection to the substrate are investigated using Raman thermography combined with device thermal simulations. This is carried out for both conventional epilayer structures, for which a method is developed to obtain more detailed temperature information through the depth of the GaN layer, and for a novel device structure which allows a spectroscopic technique to extract the temperature at a range of depths.

The theory behind both AlGaN/GaN HEMTs and the analytical techniques used in this thesis is set out in Chapter 2. Optical and related systems providing these analytical techniques have been commercially available for many years. Some principles and practicalities of their operation are described in Chapter 3. Applying these well-established measurement techniques to novel device structures not only presents new challenges, but also opens up new possibilities for understanding the operating parameters of devices. This is the subject of Chapter 4, in which layers of different compositions of AlGaN, present in the device for strain relief and improved electron confinement, are used to produce a three-dimensional temperature map of a sample. Well-established techniques may be combined in new ways. Alternatively optimising existing techniques to new samples – in this case AlGaN/GaN HEMTs with substrates transparent to visible light – can provide a source of data not previously available. This will be shown in Chapters 5 and 6. Chapter 5 demonstrates the benefits of improving the axial spatial resolution as a means

of extracting more localised temperature data from an operating device. This allows the peak temperature reached in a device to be more accurately measured; the peak temperature is particularly important for on-state device degradation. In Chapter 6, off-state degradation is investigated by electroluminescence and photoluminescence imaging and spectroscopy, combined with electrical logging. The generation of current leakage paths in the device is linked to localised light emission from the channel, and damage to the AlGaN barrier.

By using the optical methods discussed above, this thesis sets out to shed some light on the question of what limits the lifetime and performance of AlGaN/GaN HEMTs. Specifically within this overall aim, it sets out to address two broad and related questions: *"What can optimised Raman spectroscopy tell us about thermal phenomena in devices?"* and *"How can we understand more about off-state degradation processes using optical and spectroscopic techniques?"* Chapters 4 and 5 address the first question, while Chapter 6 addresses the second.

Chapter 2

Theoretical background

2.1 Materials and device structures

This section sets out the background to the material systems and devices discussed in this thesis. After an overview of the relevant history, the nitride family of semiconductors is introduced, before the electronic devices considered here are discussed.

2.1.1 Historical background

The history of GaN, at least of such a crystal quality that its electrical properties could be probed, dates back nearly 50 years, when a vapour-phase epitaxy process using ammonia was developed.[6] Early work on the use of GaN was mainly concerned with the development of a blue light emitting diode (LED) as an alternative to SiC,[13] which has an indirect bandgap,[a] a problem that was only solved over twenty years later by Nakamura.[3] GaN has a direct bandgap, making it a favourable material for photonic devices. The bandgap of GaN is around 3.4 eV, wider than most other commonly available semiconductors, which equates to a wavelength in the near ultraviolet, useful for pumping phosphors in solid state lighting, for which InGaN is also used.

At a similar time to the early optical work on GaN, work on AlGaAs/GaAs field effect transistors (FETs) led to the development of the high electron mobility transistor (HEMT). GaAs-based HEMTs are now widely used for radio frequency (RF) power amplification. The AlGaN/GaN HEMT followed soon after the GaN LED, in 1993;[5] early progress in power output was rapid, and the potential advantages of GaN-based HEMTs were soon clear.[14] In particular, the AlGaN/GaN system enables higher current densities, higher voltages and higher frequencies than AlGaAs/GaAs HEMTs, all while

[a] Which reduces the efficiency of optical devices.

operating at higher temperatures. This leads to significant performance gains in a wide range of RF applications in military, space and commercial fields, as well as the potential for high-efficiency power conversion electronics, with reduced cooling requirements compared to current (often silicon-based) technology. Despite the early developments, and ongoing improvements including the recent first deployment of a GaN amplifier on a satellite,[15] much development is still required if AlGaN/GaN HEMTs are to realise their full potential. Obtaining maximum reliability for devices operating in the most demanding applications is still a challenge, necessitating ongoing research into degradation mechanisms. For many applications, normally-off operation is desirable, and to date the majority of GaN HEMTs are normally-on, requiring a negative gate voltage to switch them off. This both increases circuit complexity (as a negative power supply may not otherwise be required and alternative circuit topographies increase component count) and leads to additional system failure modes of concern in high-power applications. A significant effort is therefore being invested in developing normally off, or enhancement mode, HEMTs, by (or funded by) major companies such as Toyota[16] and Google[17] for power conversion applications as varied as electric cars and data-centre power supplies.

The advantages of GaN as a transistor material start to become apparent when its material properties are compared to those of rival technologies. Table 2.1 compares some fundamental properties of leading semiconductor materials of interest to high-frequency/ high-power applications.

Table 2.1: High-frequency-related material parameters. Based on a table in reference 18, some data from reference 19.

	E_g [a] eV	n_i [b] cm^{-3}	ϵ_r [c]	μ_n [d] $\mathrm{cm\,V^{-1}\,s^{-1}}$	v_{sat} [e] $\times 10^7 \mathrm{cm\,s^{-1}}$	E_{br} [f] $\mathrm{MV\,cm^{-1}}$	TC [g] $\mathrm{W\,m^{-1}\,K^{-1}}$	JM [h]
Si	1.1	1.5×10^{10}	11.8	1350	1.0	0.3	150	1
GaAs	1.42	1.5×10^{5}	13.1	8500	1.0	0.6	43	2.7
SiC (4H)	3.26	8.2×10^{-9}	10	700	2.0	3.0	450 [i]	20
GaN	3.4	1.9×10^{-10}	9.0	2000 [j]	2.5	3.3	130	27.5
Diamond	5.4	1.6×10^{-27}	5.5	1900	2.7	5.6	2000	50

[a] Bandgap.
[b] Intrinsic carrier density at room temperature.
[c] Relative permittivity.
[d] Electron mobility.
[e] Saturation velocity.
[f] Breakdown field.
[g] Thermal conductivity at room temperature. The thermal conductivity reduces with temperature at typical operating temperatures,[20] for GaN $TC \propto T^{-1.4}$ is often used, though values vary.
[h] Johnson figure of merit, which compares the power-frequency performance of materials, normalised to the value for Si. $JM = {}^{E_{\mathrm{br}} v_{\mathrm{sat}}}/_{2\pi}$.
[i] In-plane; parallel to the c-axis the thermal conductivity is 330 $\mathrm{W\,m^{-1}\,K^{-1}}$.
[j] Within the 2DEG; the bulk value is 1200 $\mathrm{cm\,V^{-1}\,s^{-1}}$.

2.1.2 Crystal properties of GaN and related materials

In general GaN is grown by metal-organic chemical vapour deposition or molecular beam epitaxy. Metal-organic chemical vapour deposition is preferred for commercial devices while molecular beam epitaxy is particularly useful in research. In metal-organic chemical vapour deposition a reaction between trimethylgallium ($Ga(CH_3)_3$) and ammonia (NH_3) vapours takes place on the surface of a sample, which is held at an elevated temperature. Molecular beam epitaxy systems work at slightly lower, though still elevated, substrate temperatures. In these systems, beams of the constituent atoms are directed at the substrate, and the epilayer is slowly built up, one monolayer of material at a time in a well-controlled system. Although GaN, AlN and AlGaN crystallise in both the wurtzite and zincblende structures, for the majority of semiconductor devices constructed from this family of materials, including all those discussed here, only the wurtzite structure, shown in Figure 2.1(a) is of relevance. Growth is generally along the *c*-axis of the wurtzite crystal, normal to the surface of the substrate, with the gallium face away from the substrate and used for device growth (this is the case for all the devices discussed in this thesis), though devices have also been fabricated on the nitrogen face[21] ("N-polar") as well as on the non-polar a- and m-planes and semi-polar planes.[22]

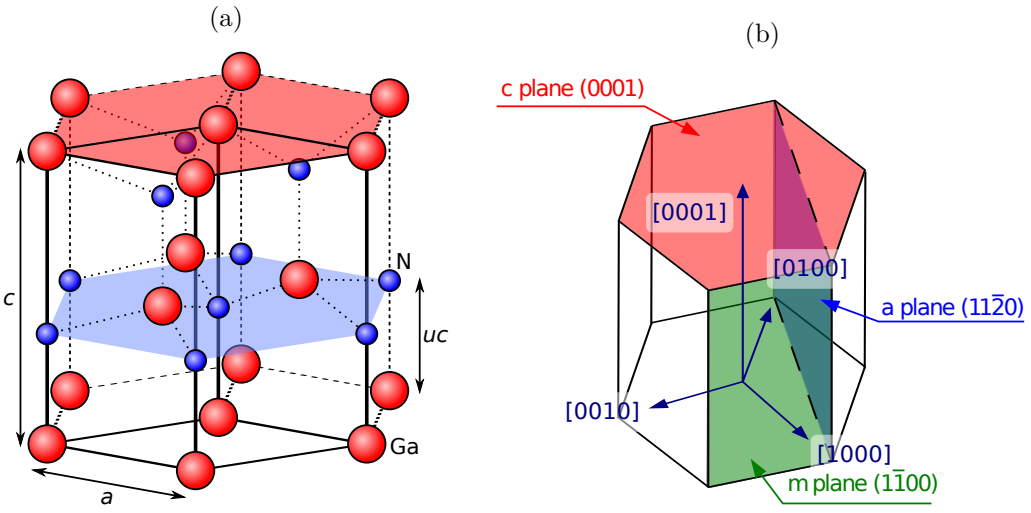


Figure 2.1: (a) The wurtzite crystal structure of GaN, showing its hexagonal nature, with the unit cell and (0001) plane highlighted. (b) The major planes in wurtzite GaN.

Figure 2.1(b) shows the orientation of the major planes used in today's GaN research. The most common plane for device growth is the c plane (0001). Wurtzite GaN belongs to the space group C_{6v}^4 ($P6_3mc$),[23] with lattice parameters $a = 3.189$ and $c = 5.185$ at room temperature. The wurtzite structure consists of two interpenetrating hexagonal close packed lattices, one for each atomic species. The internal parameter u in Figure 2.1(a) is

defined as the ratio of the Ga–N bond length to the c lattice parameter (unit cell height). Ideally $u = {}^3/_8$, in GaN it has been found experimentally to be 0.377. The (0001) face, highlighted in red in Figures 2.1(a) and 2.1(b), and the $(000\bar{1})$ face, highlighted in light blue in Figure 2.1(a), are not geometrically equivalent and are referred to as the Ga face and N face respectively. This important fact gives rise to the polarity of GaN crystals.

When considering the behaviour of electrons or phonons in a crystal, such as for Raman spectroscopy, it is useful to work in reciprocal space, or k-space, defined in terms of the reciprocal of the lattice spacing. The wave-vector $\boldsymbol{k}$ of a phonon is represented by a point in reciprocal space. The point where the reciprocal lattice vector $\boldsymbol{k} = 0$ is defined as the Γ point, the centre of the first Brillouin zone. In a wurtzite crystal the Brillouin zone takes the form of a hexagonal prism, and is depicted in Figure 2.2. In GaN, the conduction band minimum and valence band maximum are located at the Γ point, as shown in Figure 2.3. This is important because it means that GaN has a *direct* band gap, i.e. for an electron in the conduction band to annihilate with a hole in the valence band requires no phonon interaction for momentum conservation.

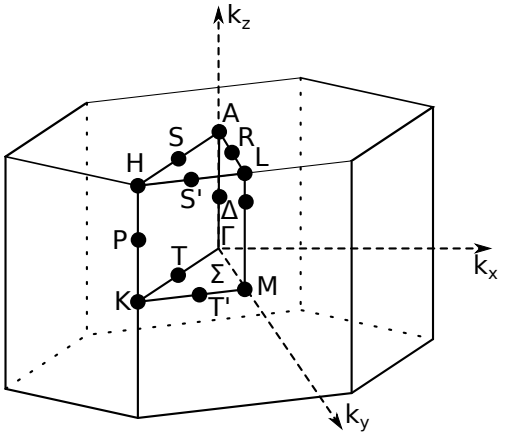


Figure 2.2: The first Brillouin zone in a wurtzite crystal, showing the points of high symmetry. Based on a figure in Reference 24.

The relationship between the reciprocal lattice vector and the energy, $E(\boldsymbol{k})$ is the band structure, which can be calculated using a variety of methods. An effective early approach was the pseudopotential method used by Bloom *et al.*[25] Later work has tended to use the linear combination of atomic orbitals[26] or the local density approximation,[27] an example of which can be seen in Figure 2.3.

Generally speaking, semiconductor *devices* rely on a difference between semiconductor *materials* for their operating principles. In many cases, for example p-n junctions, the only difference between these materials is the doping (intentional addition of specific

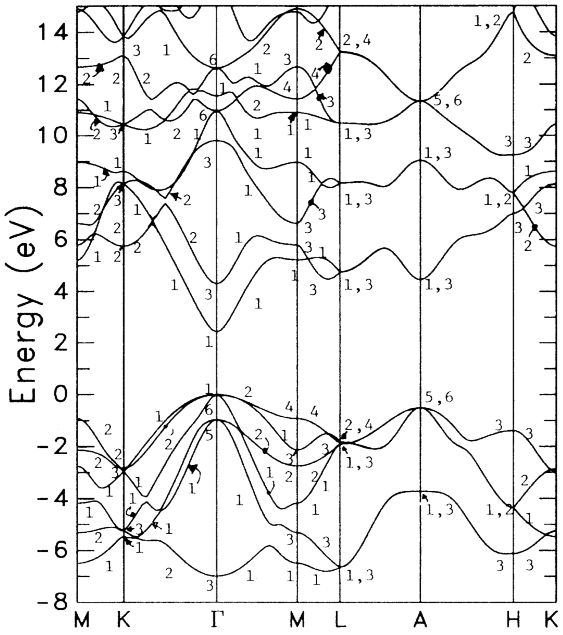


Figure 2.3: The band structure of GaN, calculated using the local density approximation by Lambrecht *et al.*, who note the need for an upward correction of the conduction band energy by a constant 0.98 eV. Figure from Reference 27.

impurities) of an otherwise identical semiconductor element or compound. In many III-nitride semiconductor devices, a heterostructure is used, i.e. the materials are different in their composition, not merely differently doped. By growing a ternary alloy consisting of aluminium (or indium) in addition to gallium and nitrogen, a range of bandgaps can be obtained. This range, for the III-nitride family of materials, is shown in Figure 2.4. It can be seen from this figure that not only is the bandgap affected by the changing composition, but so are some of the other material properties, for example the lattice constant. This fact becomes important when the epitaxial growth of heterostructures is considered, as growing an epilayer of one aluminium fraction on a layer with a different fraction will cause strain in both layers, compressive in one and tensile in the other, as the lattice constant will not change abruptly at the interface.

2.1.3 The high electron mobility transistor (HEMT)

The applications of RF systems are many and varied, with a long history driving important developments in electronics starting with vacuum tubes and the earliest transistors. Of course, many of the major developments in RF technology have been driven by military demands for communication and information, leading to major advances in first radio then radar and on to satellite data communications. In fact there is an ever growing demand for improvements in bandwidth and power in RF applications. To go

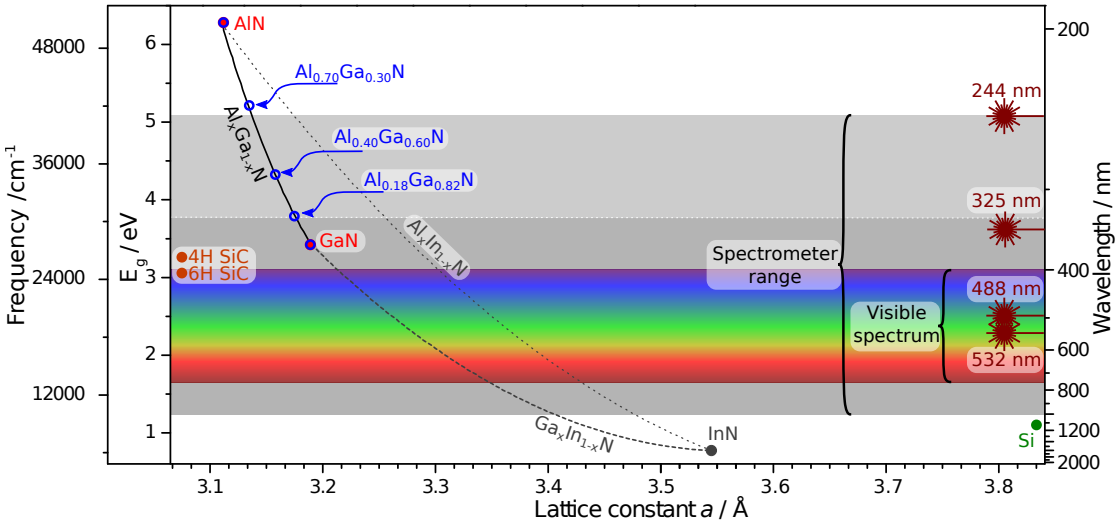


Figure 2.4: The lattice constants and bandgaps of the III-nitrides lie along the black curves – of most interest here is $Al_xGa_{1-x}N$ (solid curve). For ternary compounds the lattice constants are calculated using Vegard's law;[28,29] to calculate the bandgap an additional bowing parameter must be taken into account; these are taken from References 30, 31, and 32. GaN, AlN and some of the $Al_xGa_{1-x}N$ compositions of particular interest later in this work are highlighted, as are the important substrates SiC and Si. The laser wavelengths used for the experiments discussed later in this thesis are shown; and the relationship between eV, absolute wavenumbers (cm^{-1}) and wavelength may be seen from the vertical axes.

beyond the capabilities of the silicon transistor required new materials with properties such as higher breakdown fields and higher electron mobilities, a need initially met by GaAs, but even that technology is being left behind by the latest performance requirements, especially for high voltage and high current density operation. GaN is not the only candidate material for the latest devices: SiC transistors are commercially available, and diamond as a semiconductor is the focus of much research; however GaN has the advantage of higher electron mobility and breakdown field over SiC; diamond, though promising[33] and in some ways superior to GaN (Table 2.1) is far from the market. In addition to the RF applications, the drive for increased energy efficiency combined with new power generation systems and new applications for power conversion has led to an interest in high-performance semiconductors for use in switching power supplies. The main role that new semiconductor materials can play in these applications is to enable improved FETs, as these transistors perform switching and amplification tasks in both power supply electronics and RF systems.

A HEMT is a heterostructure field effect transistor (HFET), which is based on confining the current inside the transistor to a thin channel at the interface of two materials, forming a two-dimensional electron gas (2DEG). The structure of a HEMT is illustrated in Figure 2.5(a). It consists of a substrate (which will be discussed in more detail in

Section 2.1.4) on which is grown a nucleation layer, usually AlN, then the buffer. In most HEMTs the buffer is GaN, although a device using a series of AlGaN layers to form the buffer will be discussed in Chapter 4. Typical buffer thicknesses are no more than a few microns. On top of the buffer a much thinner barrier of AlGaN (up to a few tens of nanometres) is grown. Ohmic *source* and *drain* contacts allow electrical connection to be made to the 2DEG, resulting in a *channel* for electrons to flow from source to drain. A *gate* contact forming a Schottky junction is added on top of the buffer. The band bending illustrated in Figure 2.5(b) results in an appreciable population of electrons in the conduction band at ambient temperatures, and creates the 2DEG where the conduction band dips below the Fermi level E_F at the top of the buffer. A voltage applied to the gate contact modulates the band bending at the interface under the gate, which enables the conduction in the channel under the gate to be controlled, for example to turn off the transistor.

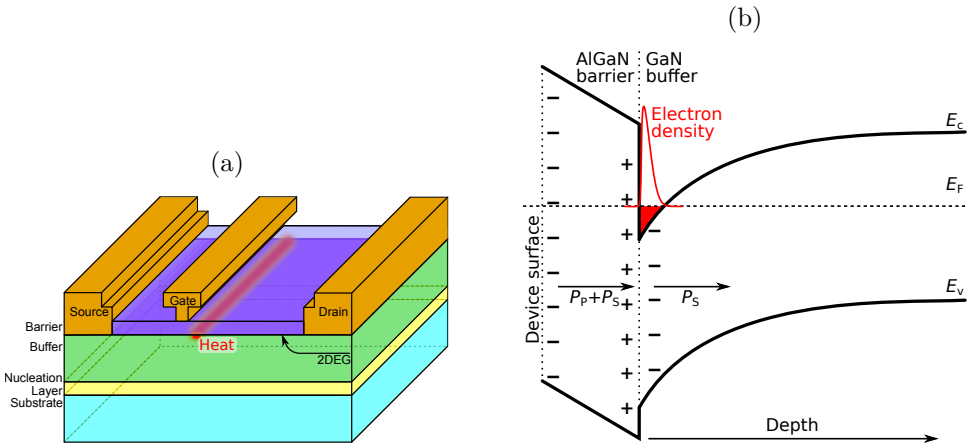


Figure 2.5: (a) The structure of a typical HEMT, consisting of a substrate on which epilayers are grown to form the nucleation layer (NL), buffer, and barrier, before contacts are deposited. The 2DEG forms at the interface between buffer and barrier, while the majority of the heat dissipation in the device, important for reliability and for much of this work, takes place near the gate contact.[34] (b) Diagram of band bending and polarisation fields at the barrier-buffer, i.e. AlGaN-GaN interface, which help to form the 2DEG.

The wurtzite crystal structure gives rise to a spontaneous polarisation charge $\boldsymbol{P}_S$ within the crystal along the [0001] direction.[35,36] In addition, wurtzite crystals are piezoelectric, leading to a further polarisation charge $\boldsymbol{P}_P$ if the crystal is strained, as is the case when one material is grown on another material with a different lattice constant such as AlGaN on GaN (Figure 2.4). The total polarisation present is therefore the sum of these two polarisations, i.e. $\boldsymbol{P} = \boldsymbol{P}_S + \boldsymbol{P}_P$ as depicted in Figure 2.5(b). The piezoelectric

polarisation is related to the strain ϵ by[36]

$$P_{\mathrm{P}i} = \sum_j e_{ij}\epsilon_j \tag{2.1}$$

where e_{ij} are the components of the polarisation tensor, and Voigt notation is used.

To calculate the polarisation which contributes to the 2DEG, it is sufficient to consider only polarisations along the [0001] axis shown in Figure 2.5(b), allowing the spontaneous polarisation to be simplified to $\boldsymbol{P}_{\mathrm{S}} = P_{\mathrm{S}}\hat{\boldsymbol{z}}$. The piezoelectric polarisation may then be expressed as

$$P_{\mathrm{P}3} = e_{33}\epsilon_3 + e_{31}(\epsilon_1 + \epsilon_2) \tag{2.2}$$

Then if a_0 and c_0 are the relaxed values of the lattice parameters a and c, $\epsilon_3 = {}^{(c-c_0)}/_{c_0}$ is the strain along the c axis and assuming the in-plane strain to be isotropic, $\epsilon_1 = \epsilon_2 = {}^{(a-a_0)}/_{a_0}$. Shear strain is ignored.

The buffer in a typical AlGaN/GaN HEMT is thick enough that strain due to heteroepitaxy may be neglected at the buffer-barrier interface; as the barrier is much thinner it will be strained due to the lattice mismatch with the buffer. The piezoelectric polarisation in the buffer may therefore be neglected, but must be considered in the barrier. This is illustrated in Figure 2.5(b) by the presence of the piezoelectric polarisation $\boldsymbol{P}_{\mathrm{S}}$ only in the barrier. This leads to a step-change in the polarisation field and therefore a net positive charge at the AlGaN side of the AlGaN/GaN interface.[37] The accumulation of electrons, which come from donor-like states at the surface of the AlGaN,[38] that this causes on the other side of the interface, in the GaN, leads to the formation of the 2DEG.

As III-nitrides are piezoelectric, the converse piezoelectric effect is relevant to devices, as it leads to a strain in response to an applied voltage. The electric fields near the gate foot can reach ~10^6 $\mathrm{V\,cm^{-1}}$, especially in the off state. An example of the electric field and the strain it causes may be seen in Figure 2.6. In many devices one or more field plates are used to cap the maximum field experienced at the gate foot, as without them the field could exceed the breakdown field of the materials used.

In the GaN system, typical HEMTs are normally-on (depletion mode), i.e. the Fermi level is above the bottom of the conduction band in the 2DEG under the gate when the gate is unbiased, as illustrated in Figure 2.5(b). A negative gate bias causes the channel to *pinch off*, i.e. stops the flow of current. This effect is depicted in Figure 2.7, where

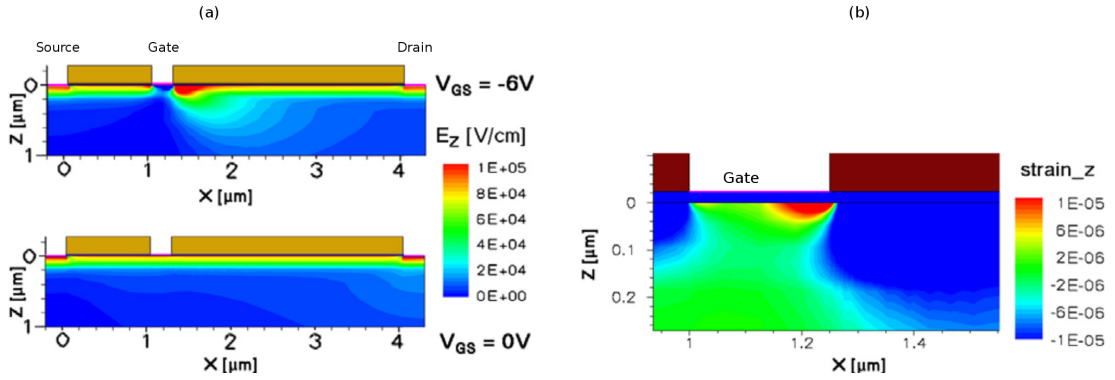


Figure 2.6: (a) The vertical electric field in a simulated HEMT at $V_{DS} = 3$ V in off- (upper) and on- (lower) states. (b) The vertical strain caused by the high field and converse piezoelectric effect in the same device in the off state ($V_{DS} = 3$ V, $V_{GS} = -6$ V). Figures from Reference 39.

the drain current I_D for a given drain-source voltage V_{DS} may be seen to reduce as an increasing negative bias is applied to the gate. Also in Figure 2.7 it may be seen that in the *linear region* the drain current is approximately linear with the drain-source voltage. In the *saturation* region self-heating limits the current by reducing the electron mobility. As the power dissipated as heat in the device is given by $P = V_{DS}I_D$, it is desirable to operate as close to the axes of Figure 2.7 as possible. A non-zero drain-source on-resistance (R_{on}) will always lead to a voltage drop when there is current flowing, shown by the position of the typical operating on state. Many reliability experiments need to dissipate more power in the device to mimic the self-heating present in HEMTs delivering significant amounts of power to a load. This is usually achieved by operating the device directly across a power supply, leading to operation in the saturation region, as shown in Figure 2.7. In a hard-switched system (for example a power supply) switching transients (dashed lines in Figure 2.7) account for significant losses. As any transistor switches it will pass through a region in which the voltage drop across it and the current flowing through it are high compared to when it is fully on or off. The time spent in this condition, and the magnitude of any overshoot, are clearly targets for reduction if power losses and self-heating are to be reduced.

Enhancement mode, "e-mode," or normally-off, devices are the focus of much development. In e-mode devices a small positive gate voltage is required for current to flow in the channel as without this voltage the Fermi level under the gate is below the conduction band. This may be achieved by a number of methods, such as fluorine implantation,[42] a p-doped GaN gate,[43] or a recessed gate.[44] Similar to a recessed gate is the use of a very thin AlGaN barrier, which must be combined with control of the (donor-like) surface states. A device of this type is studied in Chapter 4.

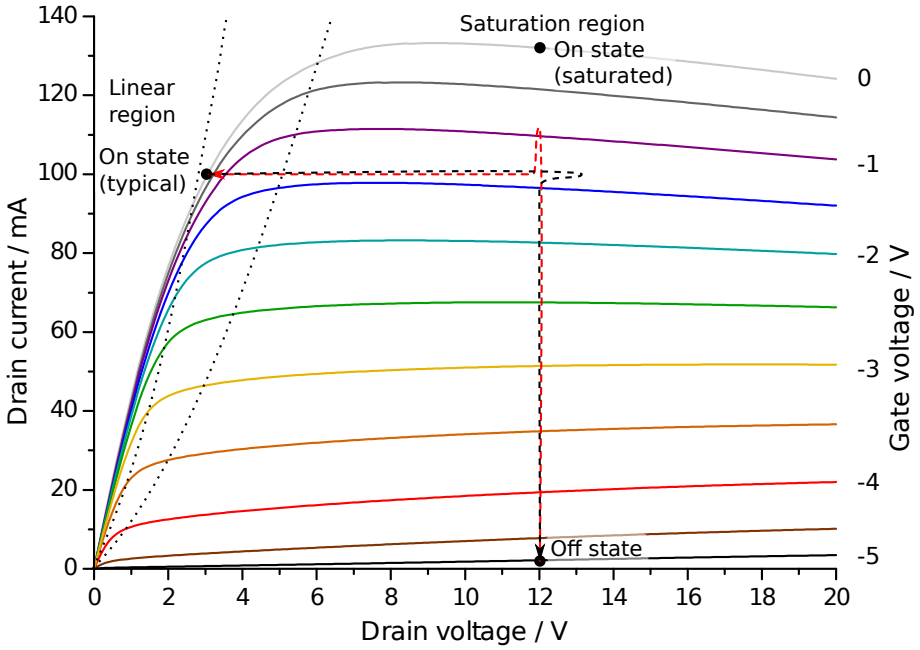


Figure 2.7: Current-voltage (I-V) curves measured on a typical depletion mode high electron mobility transistor. As the negative gate bias increases, the current flowing from source to drain decreases. Some typical operating points are shown: a pinched off state at $V_{DS} = 12$ V and $V_{GS} = -5$ V; an on state ($V_{GS} = 0$ V) in which the voltage drop across the HEMT is reduced as it is switching power into a load; and a high-power-dissipation on state in which V_{DS} is maintained at 12 V and $V_{GS} = 0$ V, used to study self-heating in the device and referred to as "on" in this thesis and much other work on device reliability. Example switching trajectories for the transition between off and the typical operating on state are illustrated by dashed lines. As the device is turned on (i.e. as V_{GS} increases towards 0 V, shown by the red dashed line), the drain current increases and may overshoot,[40] before the drain voltage drops.[41] As the device is turned off (black dashed line), the drain voltage increases, again overshooting, before the current decreases.

2.1.4 Substrates for high electron mobility transistors

As was shown in Figure 2.5(a), AlGaN/GaN HEMTs are grown on a substrate, usually of another material, as GaN is not easily grown in bulk, although devices have been grown on GaN substrates.[22] The growth of bulk GaN is the subject of much research, and is based on either (i) growing a thick GaN layer on a substrate which is removed after the GaN has been grown or (ii) ammonothermal (a slow process compared to more common growth techniques) or hydride vapour-phase epitaxial growth on a seed crystal. The choice of substrate for device growth is important for two main reasons: (i) a considerable amount of heat may be generated in the channel of a HEMT and must be removed through the substrate (therefore the substrate thermal conductivity is critical to device performance); (ii) device performance depends strongly on crystal quality, for example the threading dislocation density,[45,46] which in turn depends on the

growth conditions and substrate. Both crystal quality and heat transfer properties have implications for the reliability and performance of devices.

Sapphire (Al_2O_3) was used for much of the early work on (Al,In)GaN semiconductors; sapphire is inexpensive, insulating and convenient for LED development, being transparent to visible light. Indeed sapphire is still the most common substrate for GaN-based LEDs, with SiC used for some high-efficiency LEDs and Si proposed for lower cost, according to a recent market review.[47] However sapphire has a significant lattice mismatch (13.8%) and thermal expansion mismatch with GaN,[48] as well as poor thermal conductivity, material properties which are not conducive to obtaining the performance of which GaN-based HEMTs are capable. More recently, and more significantly for this work, the preference for substrates for AlGaN/GaN HEMTs is for silicon carbide or silicon – the devices studied in this work were all grown on silicon or silicon carbide.

Diamond used as a substrate (or intermediate heat-spreading layer) has been the focus of recent efforts due to its extremely high thermal conductivity. This usually involves growing GaN on another substrate which is removed before the GaN is bonded to the diamond. Diamond substrates are promising but improvements are still needed to realise their full potential.[49] This is partly due to the need to bond the device to the diamond, or in other designs the need to grow diamond under less-than-optimal conditions to avoid damage to the device.

Silicon is an attractive substrate for GaN devices for many reasons. Integration with silicon-based processes, such as complementary metal-oxide semiconductor (CMOS), is desirable for obtaining maximum functionality at the level of the packaged component. One example of this is integrating GaN-based HFETs with CMOS logic circuits in a single package, however the need to avoid overheating the Si logic circuit with heat from the GaN device makes thermal design a challenge. If the cost of developing manufacturing facilities for III-nitrides is to be kept within reasonable limits, the existing investment in fabrication facilities for Si suggests a requirement to use it as a substrate. Silicon wafers are also cheaper than SiC and available in larger sizes (leading to economies of scale for manufacturing). The thermal conductivity of silicon, while poor compared to SiC, is only slightly lower than that of GaN, and much higher than that of sapphire, making thermal management a significant but not impossible part of device design. Silicon as a substrate material is not without its drawbacks, however. The lattice mismatch with GaN is quite significant at 16%, comparable to sapphire, as is the thermal expansion mismatch. This

is addressed at least to some extent by strain relief layers, as will be shown in Chapter 4, these however may have an adverse effect on the thermal conductivity of the epilayer stack.

Silicon carbide is currently the substrate of choice for the highest power applications, as its thermal conductivity is much higher than that of Si. The crystal quality of GaN grown on SiC by metal-organic chemical vapour deposition has become very high,[14] but as with heteroepitaxy in general is critically dependent on the growth of an optimised nucleation layer between the substrate and the buffer.[50]

Regardless of the choice of substrate the epitaxy is not trivial. The purpose of the AlN nucleation layer is to improve the epitaxy of the buffer, including by taking up some of the lattice mismatch between substrate and buffer. Thermal effects in heterostructures affect the rate at which heat can be transported to the substrate and hence to the heatsink. A boundary between two dissimilar materials, even perfect materials, presents a resistance to heat flow,[51–53] known as a thermal boundary resistance (TBR). In a real device, especially in the presence of a lattice mismatch, the crystal quality of the epilayer is not perfect, leading to further thermal resistances such as due to phonon scattering.[54] This effect is stronger when the double interface SiC–AlN–GaN is considered, and the crystal quality of both the AlN and the GaN nearest the substrate is reduced compared to pure crystals.[55] The whole resistance across the interface may be conveniently described as an effective thermal boundary resistance,[56] while the GaN thermal conductivity may be treated as constant with depth.[57] This simplification is tested in Chapter 5.

2.1.5 Phonon modes in GaN and related materials

Lattice vibrations, or phonons, are important in semiconductors for a number of reasons. Phonons are responsible for the majority of heat transport in non-metallic crystals,[58] meaning that the thermal conductivity of semiconductors depends strongly on the scattering processes affecting phonons, such as at dislocations, impurities and material interfaces. The electrical performance is strongly affected by electron-phonon scattering, which limits the electron mobility.[59]

The phonons in a crystal may be studied by various means, among which one of the most convenient is Raman scattering. First order phonon Raman scattering takes place at the Γ point of the reciprocal lattice, and is the type of Raman scattering used in

this thesis. In a wurtzite crystal, group theory predicts eight phonon modes at the Γ point, six of which (A_1+E_1+2B_1+2E_2) are optical, and two (A_1+E_1) acoustic and therefore not accessible in Raman scattering.[60] Not all of the optical modes are Raman-active however – the two B_1 modes are silent, i.e. they are visible in neither Raman nor infrared measurements. The E_1 and E_2 phonons are degenerate, and the A_1 and E_1 optical phonons shift in frequency depending on whether the mode is longitudinal (LO) or transverse (TO).[61] Second order Raman scattering may be used to investigate the phonon frequencies at other points in the reciprocal lattice.[23] The lattice vibrations making up the Raman-active modes in wurtzite crystals are shown in Figure 2.8(a), while the phonon dispersion curve (i.e. the relationship between the lattice points and the phonon frequency) for GaN is shown in Figure 2.8(b).

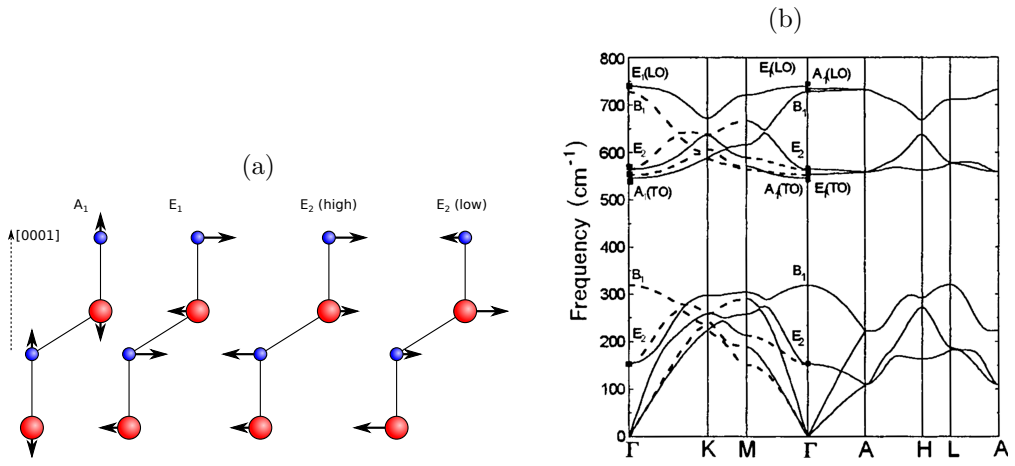


Figure 2.8: (a) The Raman-active phonon modes in wurtzite crystals such as GaN.[62] (b) The phonon wavevector-energy dispersion curves for GaN, taken from Reference 23.

2.2 Raman spectroscopy and thermography

In this section Raman scattering in general, and particularly in its application to the thermography of semiconductors, is introduced. Although Raman scattering has a history dating back to the 1920s,[63] it is only since the development of lasers that it has had a useful analytical role. The aim of Raman thermography is to measure the temperature of a sample by means of measuring the change in frequency with temperature of the sample's characteristic spectral lines. Figure 2.9 illustrates this process and hints at some of the challenges involved. The x-axis in Figure 2.9 is denominated in wavenumbers (units of frequency widely used in spectroscopy, where 1 cm^{-1}= 2.998×10^{10} Hz). Raman band positions are usually expressed as an offset or *shift* from the laser frequency

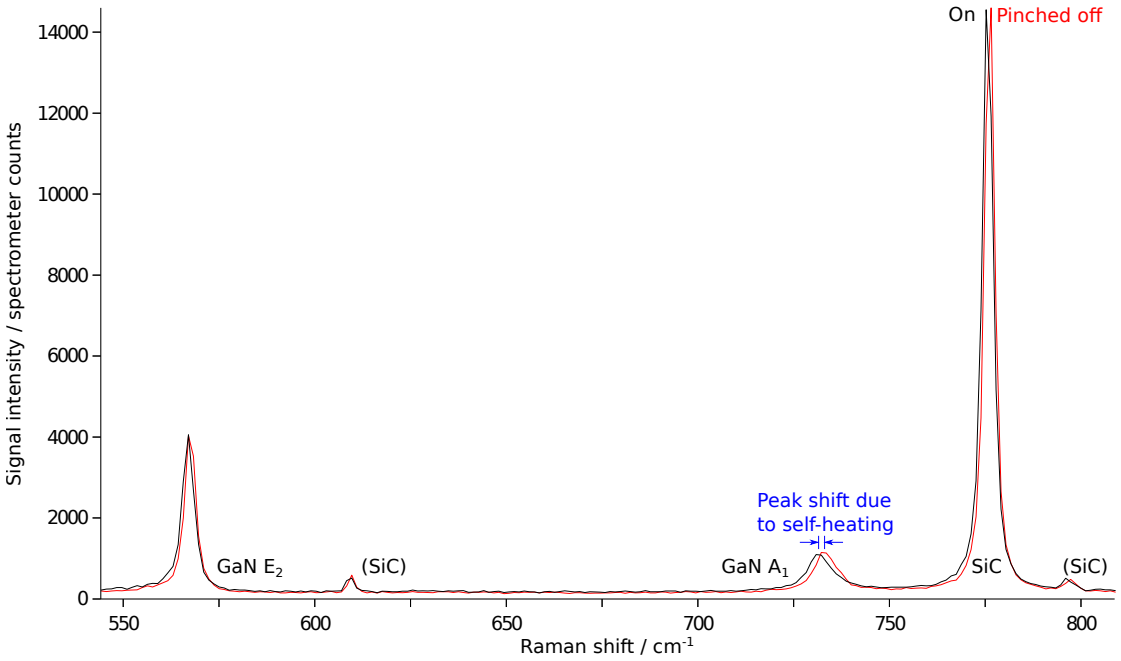


Figure 2.9: Two spectra from an AlGaN/GaN HEMT on a SiC substrate at $V_{DS} = 30$ V: Pinched off ($V_{GS} = -7.8$ V, red curve) and on ($V_{GS} = 0$ V, $I_D = 43$ mA, black curve). Two clear phonon lines from the GaN buffer and one from the substrate are visible, along with weaker substrate-related features.

(a 488 nm Ar-ion laser in the case of Figure 2.9 and much of the work described in this thesis). Wavenumber shift is used because it is characteristic of the phonon, i.e. independent of the laser frequency. In Raman thermography it is necessary to precisely locate the peak frequencies, as the shift with temperature (illustrated for the GaN A_1 line in Figure 2.9) is rather small.

When a photon of light encounters a material it may be scattered. The vast majority of scattered photons are scattered elastically – Rayleigh scattering – that is with no change in energy. Other scattering processes are inelastic, involving an exchange of energy with the material, and are generally less likely; these include Raman scattering and photoluminescence. Rayleigh scattering, Raman scattering and fluorescence (or photoluminescence) are illustrated in Figure 2.10. On the left of Figure 2.10, Rayleigh scattering is depicted, in which the re-emitted photon has the same energy as the absorbed photon.

In Raman scattering, the incident photon may be regarded as exciting a virtual state in the material, which then decays leaving a change in vibrational energy. When the scattered photon has a lower energy than the incident photon this is known as "Stokes" scattering – a term acquired from fluorescence; an emitted photon at higher energy is described as "anti-Stokes". Stokes and anti-Stokes peaks occur symmetrically about the laser line, at $\omega_{\text{Stokes}} = \omega_{\text{Laser}} - \omega$ and $\omega_{\text{anti-Stokes}} = \omega_{\text{Laser}} + \omega$, where ω_{Stokes} and

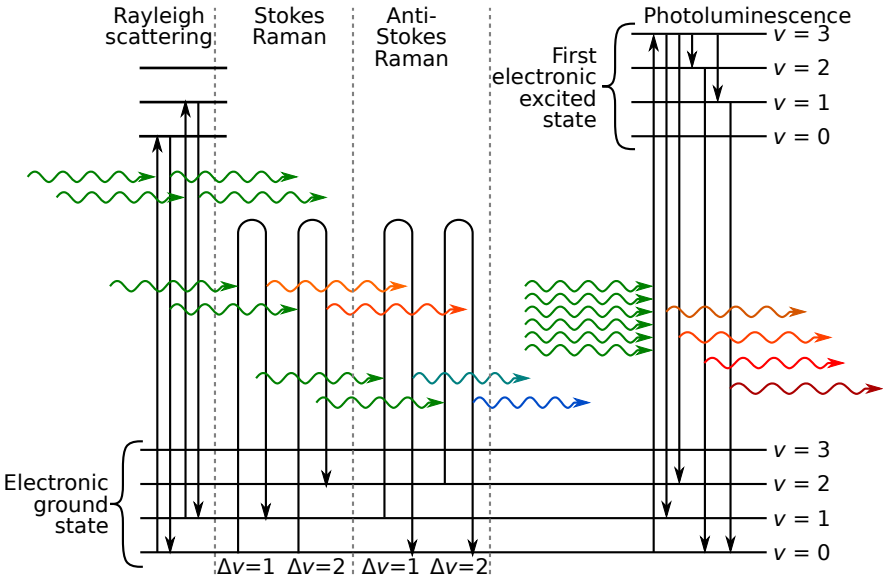


Figure 2.10: Some major photon-scattering processes. Left to right: Rayleigh (elastic) scattering, Raman (inelastic) scattering, and photoluminescence (absorption, relaxation and re-emission). Vibrational energy levels are indicated as v. Based on figures from References 63 and 64.

$\omega_{\text{anti-Stokes}}$ are the frequencies of the Raman-shifted light, ω_{Laser} is the laser frequency and ω is the frequency of the vibrational transition, as illustrated in Figure 2.11. The relative intensity of the Stokes and anti-Stokes peaks is determined by the occupancy of a raised vibrational energy levels ($v > 0$ in Figure 2.10) in the sample, and therefore the sample temperature.

For many combinations of material and laser wavelength photoluminescence is a much more likely process, and is shown to the right of Figure 2.10. In photoluminescence, a photon is absorbed causing an electronic excitation, which decays via one or more relaxation processes before emitting a photoluminescence photon.

The Raman effect is particularly weak (only 1 in 10^6 to 10^8 incident photons is Raman-scattered) so an intense monochromatic light source is needed to obtain useful information. Historically, important results were obtained by using mercury arc lamps, but they have been completely replaced by lasers in modern Raman spectroscopy. In addition, before the widespread availability of charge-coupled devices (CCDs), obtaining Raman spectra was a time-consuming business carried out in complete darkness as the scattered light was collected by a photomultiplier tube (PMT), photographic film, or a photodiode array. More recently the development of high-throughput spectrometers, using sophisticated holographic or dielectric thin-film filters to reject the unwanted Rayleigh scattered light, while not appreciably attenuating nearby frequencies, has allowed Raman spectroscopy to reach into new areas. Formerly this task required multiple diffraction gratings

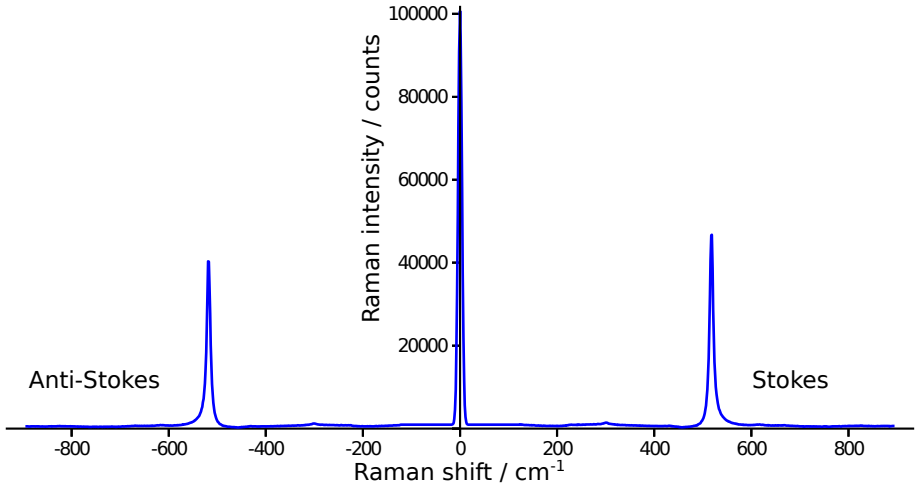


Figure 2.11: Illustrative Raman spectrum of Si at room temperature, showing the laser line at 0 cm^{-1}, and the Si peaks at ± 520 cm^{-1}.

(which are still used for measuring Raman bands within a few wavenumbers of the exciting line), the use of sampling geometries which minimised the amount of laser light reaching the spectrometer, or commonly both.

While Raman spectroscopy can be used for material identification, and has been applied to fields as diverse as historical forensics[65] and checking the contents of tablets for pharmaceutical applications,[66] in the present work it is known which materials are present and where, the task instead is to probe the temperature (and possibly strain) within the sample.

2.2.1 Porto notation

It is sometimes necessary in Raman spectroscopy to consider the polarisation and propagation of incident and scattered light with respect to a suitable reference frame (such as the crystal axes). A convenient way of describing these directions was proposed by Porto, and is named after him.[67,68] The geometry of an experiment is described by

$$i(jk)l \tag{2.3}$$

where j and k are the polarisation directions of the incident and scattered light respectively, and i and l are the propagation directions, again of the incident and scattered light respectively.

As an example, a Raman microscope system (such as those used in this work) generally uses the backscatter geometry, in which i and l are antiparallel. If the direction of the

incident light is taken to be the z axis, the backscattered light is deemed to propagate along the $\bar{z}$ axis. If a linearly polarised laser is used, as is common, the polarisation specified by j in Equation 2.3 may be assigned to the x direction. An analyser fitted to the spectrometer causing it to be sensitive to linear polarisation parallel to the incident light will also be in the x direction; a perpendicular analyser would be in the y direction, these values being assigned to k. When no analyser is used, and the spectrometer is sensitive to all incoming polarisations, k may be denoted by a dash $(-)$. Thus the Porto notation used in the majority of this work is described by $z(x-)\bar{z}$, however for some experiments in this work the laser polarisation was modified, which will be discussed in Section 2.3.5 and Chapter 5.

When working with crystalline materials, the spectrometer axes are generally aligned to the crystal axes of the sample, for example when studying epilayers grown along the c crystal axis, the z experimental axis may be set parallel to the c crystal axis, which is the case for the typical c-plane HEMTs studied here. For typical measurements in this thesis, given the crystal structures of most of the materials used in AlGaN/GaN HEMTs, it is not necessary to align the polarisation to the axes of the crystal, instead the device structure is typically aligned to the spectrometer input.

2.2.2 Classical theory of Raman scattering in crystalline materials

A classical description of Raman scattering first requires an explanation of the scattering of light by matter, in the form of Rayleigh scattering.[64,69] The classical description of light scattering assumes that matter consists of dipoles bound by unspecified, but not electrostatic, forces. An electromagnetic wave, of a frequency that does not match a natural frequency of the dipoles within the medium, encountering such a material causes the dipole separation to oscillate, and the dipole radiates at the oscillation frequency. Both Rayleigh and Raman scattering may be described by considering a polarisability α, which connects a periodic applied electric field $E = a\cos(\omega_{\text{Laser}}t)$ such as in a laser beam, to a resulting dipole μ by

$$\mu = \alpha E = \alpha A \cos(\omega_{\text{Laser}}t) \tag{2.4}$$

It is important to note that the polarisability α itself depends on frequency, and is related to the electric susceptibility χ and the permittivity of free space ϵ_0 by the Clausius-Mossotti relation

$$\frac{\epsilon_0 \chi}{\epsilon_0(\chi+3)}\frac{M}{d} = \frac{N_\mathrm{A}\alpha}{3\epsilon_0} \tag{2.5}$$

where M is the molar mass of the material and d is its density, and N_A is Avogadro's number. Rayleigh scattering has an intensity proportional to the mean square of the dipole moment μ (i.e. an amplitude proportional to μ and therefore the polarisability α). If the dipole is itself vibrating (for example due to thermal excitation) with amplitude B, the polarisability will be affected, becoming

$$\alpha' = \alpha\left[1 + B\cos\left(\omega t\right)\right] \tag{2.6}$$

where ω is the angular vibrational frequency.

The electric susceptibility χ is a symmetric second rank tensor (though it reduces to a scalar in an isotropic medium) therefore the polarisability α is also a tensor in the general case. In crystalline materials, such as those used in the semiconductor devices discussed in this thesis, instead of the molecular dipole vibration used in the simple classical Raman theory, the incident light interacts with phonons – lattice vibration quanta – of frequency ω which are created (for Stokes shift) or destroyed (for anti-Stokes shift). When considering Raman scattering inside a crystalline material, Equation 2.4 may be rewritten in terms of a vector electromagnetic field $\boldsymbol{E}(\boldsymbol{r},t) = \boldsymbol{E}(\boldsymbol{k},\omega_\mathrm{Laser})\cos(\boldsymbol{k}\cdot\boldsymbol{r} - \omega_\mathrm{Laser}t)$ which varies with time t and space $\boldsymbol{r}$ giving[70]

$$\boldsymbol{\mu}(\boldsymbol{k},\omega_\mathrm{Laser})\cos(\boldsymbol{k}\cdot\boldsymbol{r} - \omega_\mathrm{Laser}t) = \chi(\boldsymbol{k},\omega_\mathrm{Laser},\boldsymbol{Q})\boldsymbol{E}(\boldsymbol{k},\omega_\mathrm{Laser}) \tag{2.7}$$

where $\boldsymbol{k}$ is the wavevector of the incident field and $\boldsymbol{Q}$ is the lattice displacement, which is related to the phonon wavevector $\boldsymbol{q}$ by

$$\boldsymbol{Q}(\boldsymbol{r},t) = \boldsymbol{Q}(\boldsymbol{q},\omega)\cos(\boldsymbol{q}\cdot\boldsymbol{r} - \omega t) \tag{2.8}$$

The adiabatic approximation is made, which assumes that the electronic frequencies which determine χ are much greater than ω, and χ is therefore taken to be a function of $\boldsymbol{Q}$. As (normally) these vibrations are small with respect to the lattice constants, it is appropriate to expand χ as a Taylor series in $\boldsymbol{Q}$, where χ_0 is the susceptibility in the

absence of vibrations

$$\chi(\boldsymbol{k}_{\text{Laser}}, \omega_{\text{Laser}}, \boldsymbol{Q}) = \chi_0(\boldsymbol{k}_{\text{Laser}}, \omega_{\text{Laser}}) + \left(\frac{\partial \chi}{\partial \boldsymbol{Q}}\right)_0 \boldsymbol{Q}(\boldsymbol{r}, t) \tag{2.9}$$

The lattice wave $\boldsymbol{Q}(\boldsymbol{r}, t)$ introduced in Equation 2.8 causes the oscillating susceptibility in the second term.

Equations 2.9 and 2.7 may then be combined to give

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{r}, t, \boldsymbol{Q}) &= \boldsymbol{\mu}_0(\boldsymbol{r}, t) + \boldsymbol{\mu}_{\text{ind}}(\boldsymbol{r}, t, \boldsymbol{Q}) \\ &= \chi_0(\boldsymbol{k}_{\text{Laser}}, \omega_{\text{Laser}}) \boldsymbol{E}_{\text{Laser}}(\boldsymbol{k}_{\text{Laser}}, \omega_{\text{Laser}} \cos(\boldsymbol{k}_{\text{Laser}} \cdot \boldsymbol{r} - \omega_{\text{Laser}} t) \\ &+ \left(\frac{\partial \chi}{\partial \boldsymbol{Q}}\right)_0 \boldsymbol{Q}(\boldsymbol{r}, t) \boldsymbol{E}_{\text{Laser}}(\boldsymbol{k}_{\text{Laser}}, \omega_{\text{Laser}}) \cos(\boldsymbol{k}_{\text{Laser}} \cdot \boldsymbol{r} - \omega_{\text{Laser}} t)\end{aligned} \tag{2.10}$$

where the first term, $\boldsymbol{\mu}_0$, is a polarisation oscillating in phase with the incident radiation and the second term, $\boldsymbol{\mu}_{\text{ind}}$ is a polarisation wave induced by a phonon. This induced polarisation may be rewritten as

$$\boldsymbol{\mu}_{\text{ind}}(\boldsymbol{r}, t, \boldsymbol{Q}) = \left(\frac{\partial \chi}{\partial \boldsymbol{Q}}\right)_0 \boldsymbol{Q}(\boldsymbol{q}, \omega) \cos(\boldsymbol{q} \cdot \boldsymbol{r} - \omega t) \times \boldsymbol{E}_{\text{Laser}}(\boldsymbol{k}_{\text{Laser}}, \omega_{\text{Laser}}) \cos(\boldsymbol{k}_{\text{Laser}} \cdot \boldsymbol{r} - \omega_{\text{Laser}} t) \tag{2.11}$$

Applying the trigonometric relation $\cos\theta \cos\varphi = {}^1\!/_2 \left[\cos(\theta - \varphi) + \cos(\theta + \varphi)\right]$ to Equation 2.11 gives

$$\begin{aligned}\boldsymbol{\mu}_{\text{ind}}(\boldsymbol{r}, t, \boldsymbol{Q}) = \frac{1}{2} \left(\frac{\partial \chi}{\partial \boldsymbol{Q}}\right)_0 \boldsymbol{Q}(\boldsymbol{q}, \omega) \boldsymbol{E}_{\text{Laser}}(\boldsymbol{k}_{\text{Laser}}, \omega_{\text{Laser}}) \times \\ \left(\cos\left[(\boldsymbol{k}_{\text{Laser}} + \boldsymbol{q}) \cdot \boldsymbol{r} - (\omega_{\text{Laser}} + \omega) t\right] + \cos\left[(\boldsymbol{k}_{\text{Laser}} - \boldsymbol{q}) \cdot \boldsymbol{r} - (\omega_{\text{Laser}} - \omega) t\right]\right)\end{aligned} \tag{2.12}$$

demonstrating that $\boldsymbol{\mu}_{\text{ind}}$ consists of Stokes-shifted and anti-Stokes-shifted parts, with frequencies symmetrically shifted about the incident frequency. These two beat frequencies at $\omega_{\text{Laser}} + \omega$ (anti-Stokes) and $\omega_{\text{Laser}} - \omega$ (Stokes) appear, albeit weakly, in the spectrum of scattered light as Raman scattering, in addition to the Rayleigh scattering at the incident frequency ω_{Laser}, which is due to the polarisation $\boldsymbol{\mu}_0$.

For a scattered photon to gain energy in anti-Stokes scattering implies that there is a vibrational excitation present from which energy may be acquired by the photon. The occupancy of excited states is temperature dependent, with an occupancy $n(\omega) = [\mathrm{e}^{(\hbar\omega/k_{\text{B}}T)} - 1]^{-1}$ for a vibrational frequency ω and temperature T, where k_{B} and $\hbar$ are, as usual, Boltzmann's constant and the reduced Planck constant respectively. Therefore

the anti-Stokes intensity, or in practice the ratio of Stokes to anti-Stokes intensities may be used as a probe of temperature.[71,72] The Stokes-anti-Stokes scattering rate ratio is given by

$$R_{\mathrm{S}/\mathrm{a\text{-}S}} = \frac{A_{\mathrm{s}}}{A_{\mathrm{a\text{-}S}}} \mathrm{e}^{(\hbar\omega/k_{\mathrm{B}}T)} \tag{2.13}$$

where the Stokes and anti-Stokes amplitudes A_{s} and $A_{\mathrm{a\text{-}S}}$ incorporate the Raman susceptibility (which depends on the wavelength λ as well as in principle the temperature T) and any other optical parameters. The dependence of A_{s} and $A_{\mathrm{a\text{-}S}}$ on T may be neglected if the excitation frequency ω_{Laser} is far from a resonant frequency of the sample, making the Stokes-anti-Stokes intensity ratio a straightforward temperature probe. However obtaining good signal-to-noise ratio for the anti-Stokes peak at moderate temperatures requires either a long acquisition time or a strong signal, especially for bands further from the laser line, where more energy must be available, so this technique is rarely used.[56]

One-phonon Raman scattering, as described above, conserves both frequency and wavevector, meaning that the wavevectors of phonons studied by one-phonon Raman scattering are limited to less than twice the photon wavevector.[70] For typical lasers and refractive indices, $|\boldsymbol{q}| \sim 10^6$ cm^{-1}, or of the order of $^1/_{100}$ the Brillouin zone size, with the result that one-phonon scattering is only sensitive to zone-centre phonons, and $\boldsymbol{q}$ can be taken to be zero.

2.2.3 Quantum theory of Raman scattering

Although the classical theory will be revisited for consideration of the Raman selection rules, it is first necessary to describe the quantum processes inherent in Raman scattering, concentrating on Stokes scattering. A quantum description of inelastic light scattering by phonons requires that the state of three systems is specified: the incident (frequency ω_{Laser}) and scattered (frequency ω_{Stokes} or $\omega_{\mathrm{anti\text{-}Stokes}}$) photons; the relevant phonon with frequency $\omega = \omega_{\mathrm{Laser}} - \omega_{\mathrm{Stokes}}$; and the electrons in the semiconductor.[70] Initially there are $N(\omega_{\mathrm{Laser}})$ photons at the laser frequency, and $N(\omega_{\mathrm{Stokes}})$ at the scattered frequency. At a nonzero temperature T there are N_q phonons present, but all electrons are assumed to be in the valence band.

After Stokes scattering, N_q and $N(\omega_{\mathrm{Stokes}})$ have increased by one, while $N(\omega_{\mathrm{Laser}})$ has decreased by one, with no effect on the electrons. Although the electronic state is the

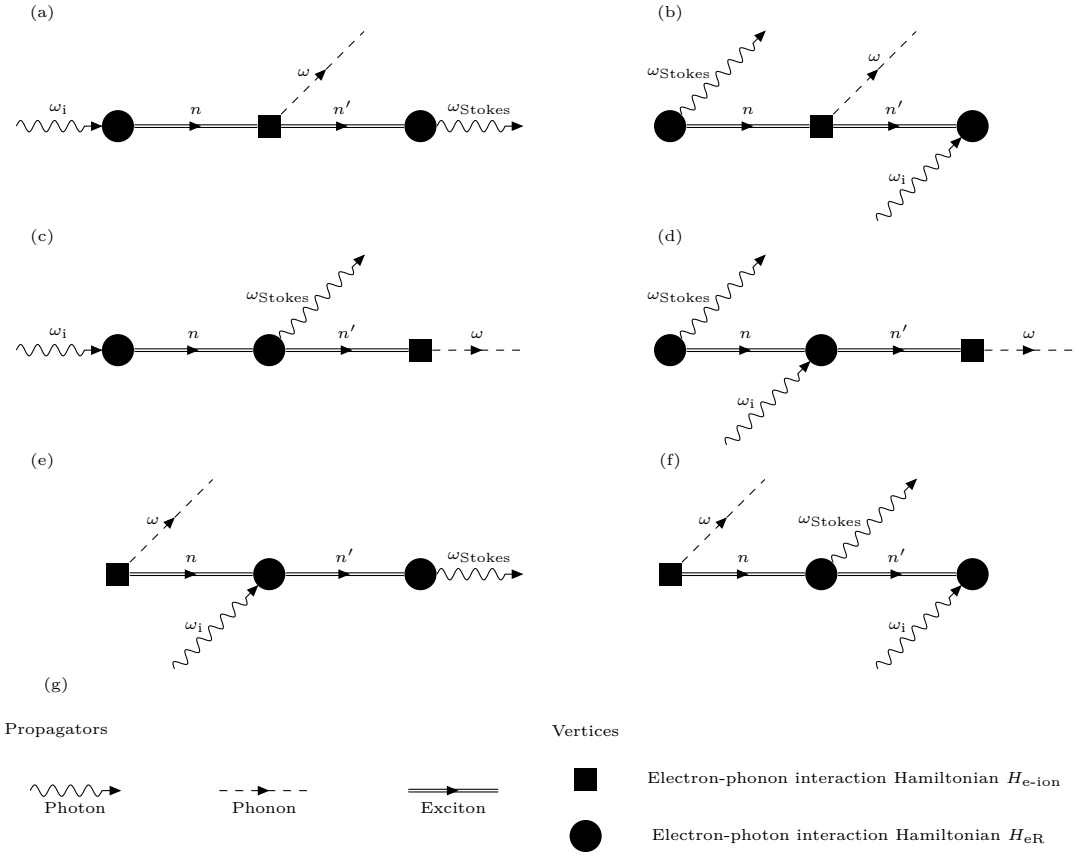


Figure 2.12: Feynman diagrams for the Stokes Raman scattering processes: (a) see main text for derivation of scattering probability; (b-f) the other possible time-sequences; (g) the notation used in the diagrams. The incident photon has frequency ω_{i}, the phonon has frequency ω and the Stokes-shifted photon has frequency ω_{Stokes}, while n and n' are intermediate virtual states. Based on a figure in Reference 70.

same before and after the scattering event, it is necessary to include the effect of electrons on the coupling of photons to phonons via the electron-radiation interaction Hamiltonian H_{eR}, as the direct interaction between photons and phonons is very weak for the large difference between ω_{Laser} and ω in Raman scattering with visible light. The scattering then takes place as follows: first the incident photon causes an electronic excitation; then this exciton scatters into another state emitting a phonon via the electron-phonon interaction Hamiltonian $H_{\text{e-ion}}$; finally the exciton recombines emitting the Raman-scattered photon. This is shown in Figure 2.12(a). The virtual transitions involving the electrons do not have to individually conserve energy – meaning that sub-bandgap illumination can still lead to Raman-scattering – but do conserve wavevectors. Energy is of course conserved over the process as a whole. This leads to further permutations for the time sequence of the scattering process, as shown in Figure 2.12(b–f).

The spontaneous Raman scattering in this process will be described first in terms of stimulated Raman scattering, and then the probabilities converted to those for spontaneous

scattering, following the approach in Reference 70.

Considering the first vertex in Figure 2.12(a), for an initial state $|\mathrm{i}\rangle$ with energy E_i and an intermediate electronic state $|n\rangle$ with energy E_n, if an energy quantum $\hbar\omega_\mathrm{i}$ is absorbed, a term

$$\sum_n \frac{\langle n|H_\mathrm{eR}(\omega_i)|\mathrm{i}\rangle}{[\hbar\omega_\mathrm{i} - (E_n - E_\mathrm{i})]} \tag{2.14}$$

is introduced into the scattering probability.

A similar term is introduced by the second vertex, with intermediate state $|n'\rangle$. In this case a quantum of energy $\hbar\omega$ is emitted (leading to the negative sign) in the form of a phonon. After simplifying the denominator the resulting probability can be written as

$$\sum_{n,n'} \frac{\langle n'|H_\mathrm{e-ion}(\omega)|n\rangle\,\langle n|H_\mathrm{eR}(\omega_i)|\mathrm{i}\rangle}{[\hbar\omega_\mathrm{i} - (E_n - E_\mathrm{i})]\,[\hbar\omega_\mathrm{i} - \hbar\omega - (E_{n'} - E_\mathrm{i})]} \tag{2.15}$$

Continuing until the last vertex is included, and taking into account that the final electronic state $|\mathrm{f}\rangle$ is identical to the initial state $|\mathrm{i}\rangle$ allows the scattering probability to be calculated according to Fermi's Golden Rule. This rule states that the scattering rate for a transition from an initial state $|\mathrm{i}\rangle$ to a final state $|\mathrm{f}\rangle$ due to a Hamiltonian H which oscillates at a frequency ω, where $|\mathrm{i}\rangle$ and $|\mathrm{f}\rangle$ differ by energy $\hbar\omega$ is given by

$$T_{\mathrm{i}\to\mathrm{f}} = \frac{2\pi}{\hbar}|\langle\mathrm{f}|H|\mathrm{i}\rangle|^2\rho \tag{2.16}$$

where ρ is the density of final states.

Because energy is conserved in the entire process, the final scattering event must include the delta function $\delta\left(\hbar\omega_\mathrm{i} - \hbar\omega - \hbar\omega_\mathrm{Stokes}\right)$. The scattering probability due to the process in Figure 2.12(a) then becomes

$$P_\mathrm{Stokes} = \frac{2\pi}{\hbar}\left|\sum_{n,n'} \frac{\langle\mathrm{i}|H_{eR}(\omega_\mathrm{Stokes})|n'\rangle\,\langle n'|H_\mathrm{e-ion}|n\rangle\,\langle n|H_\mathrm{eR}(\omega_\mathrm{i})|i\rangle}{[\hbar\omega_\mathrm{i} - (E_n - E_\mathrm{i})]\,[\hbar\omega_\mathrm{i} - \hbar\omega - (E_{n'} - E_\mathrm{i})]}\right|^2 \delta(\hbar\omega_\mathrm{i} - \hbar\omega - \hbar\omega_\mathrm{Stokes}) \tag{2.17}$$

The probabilities due to the six possible scattering sequences shown in Figure 2.12 must be summed to give the total scattering probability, then Einstein's A and B coefficients must be used to convert the probability to that of spontaneous emission.[70] A_{nm} is Einstein's coefficient for the rate of spontaneous emission of radiation due to a transition

from level n to level m, and is given by

$$A_{nm} = \frac{8\pi h\nu^3 n^3}{c^3} B_{nm} \tag{2.18}$$

where h is Planck's constant, n is the refractive index of the medium, ν is the phonon frequency (i.e. $\nu = {}^{\omega}/_{(2\pi)}$, and c the speed of light in vacuum. B_{nm} is the stimulated emission probability, and B_{mn} is the absorption probability, where $B_{nm} = B_{mn}$.

These relations theoretically allow the spontaneous Raman scattering rate to be found from material properties. Due to many unknown quantities it is however rarely possible to predict Raman scattering rates from the scattering probabilities such as those in Equation 2.17.

2.2.4 The Raman tensor and selection rules

Returning to the classical description of Raman scattering in Section 2.2.2, and considering the relationship between $\boldsymbol{\mu}$ and $\boldsymbol{E}$ given by Equation 2.12 allows the Stokes scattering intensity I_{Stokes} to be written as[70]

$$I_{\text{Stokes}} \propto \left| \boldsymbol{e}_{\text{Laser}} \cdot \left(\frac{\partial \chi}{\partial \boldsymbol{Q}} \right)_0 \boldsymbol{Q}(\omega) \cdot \boldsymbol{e}_{\text{Stokes}} \right|^2 \tag{2.19}$$

where $\boldsymbol{e}_{\text{Laser}}$ is the polarisation of the incident (laser) radiation, and the phonon wavevector $\boldsymbol{q}$ is approximated as zero (one-phonon scattering).

This classical result requires a vibration to be present, as the scattered intensity depends on the square of the vibration amplitude $\boldsymbol{Q}$; the equivalent quantum formulation would give a dependency on the phonon occupancy.

If $\boldsymbol{Q}$ is taken to be the vector displacement of an individual atom due to a phonon, ${}^{\partial\chi}/_{\partial \boldsymbol{Q}}$ will be a third-rank tensor.[70] A unit vector $\hat{\boldsymbol{Q}} = {}^{\boldsymbol{Q}}/_{|\boldsymbol{Q}|}$ may be defined, parallel to the phonon-induced displacement, allowing a second-rank tensor R, the *Raman tensor* to be defined by

$$R = \left(\frac{\partial \chi}{\partial \boldsymbol{Q}} \right)_0 \hat{\boldsymbol{Q}}(\omega) \tag{2.20}$$

The scattering intensity is now defined in terms of this Raman tensor as

$$I_{\text{Stokes}} \propto \left| \boldsymbol{e}_{\text{Laser}} \cdot R \cdot \boldsymbol{e}_{\text{Stokes}} \right|^2 \tag{2.21}$$

While group theory allows some elements of the Raman tensor for a particular phonon mode to be set to zero,[61] finding the values of the remaining terms, which may in principle be complex, requires an experimental approach. Selection rules may be derived from the Raman tensor and related to experiments. These rules relate the polarisation direction of the incoming and scattered radiation to the crystal axes, and are therefore important for the design and analysis of Raman experiments. The effect on the polarisation and emission direction of the Raman-scattered light can be described using this tensor formulation. A tensor component R_{ij} couples incident light polarised in the j direction to a vibration in the i direction, thus the emitted Raman-shifted light is emitted following the rules for a dipole in the i direction. This means that the intensity is maximised perpendicular to i, and polarised parallel to i. The Raman tensors for the Raman-active phonon modes in wurtzite GaN are as follows[73]

$$R_{\mathrm{A}_1} = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$$

$$R_{\mathrm{E}_1(x)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & c \\ 0 & c & 0 \end{bmatrix} \qquad R_{\mathrm{E}_1(y)} = \begin{bmatrix} 0 & 0 & -c \\ 0 & 0 & 0 \\ -c & 0 & 0 \end{bmatrix}$$

$$R_{\mathrm{E}_2(x)} = \begin{bmatrix} 0 & d & 0 \\ d & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad R_{\mathrm{E}_2(y)} = \begin{bmatrix} d & 0 & 0 \\ 0 & -d & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.22}$$

where a, b, c and d are elements of the Raman tensor.

As was mentioned in the discussion of phonon modes, not all modes are visible in Raman spectroscopy, but there are further constraints imposed by the accessibility of certain modes to certain polarisations as well as the need for a reasonable signal strength. From Equation 2.22 it can be seen that the E_1 mode is not accessible in $z(x-)\bar{z}$ configuration: the only components in the Raman tensors for the E_1 mode are xz, yz, zx and zy, meaning that either the polarisation of the incident light or the dipole must be aligned to z. The incident polarisation is not aligned to z for conventional polarisations of the beam for incident light propagating along the z-axis, and if the dipole is parallel to z the emission intensity is zero along z. Chapter 5 will explore attempts to work round this limitation. The E_2 and A_1 modes both have Raman tensor elements in the xy plane, and

are therefore easily observed in a $z(x-)\bar{z}$ backscatter experiment The A_1(TO) mode is also not accessible in $z(x-)\bar{z}$ orientation.[74]

2.2.5 Raman thermography

The purpose of Raman thermography is to use Raman spectroscopy to measure the temperature of a material such as, in this thesis, a component of a semiconductor device. While temperature can be measured in many other ways, some of which are similarly non-contact, such as infrared[75] and photoluminescence,[76] Raman thermography is advantageous for its high spatial resolution – around 0.5 µm – compared to both infrared and common photoluminescence configurations for wide-bandgap systems. This is important given the length scales present in HEMTs, where the source-drain distance is typically in the region of 1 µm to 10 µm.

In contrast to the Stokes-anti-Stokes approach given in Equation 2.13, temperature (and stress) in crystalline materials can be measured by studying the interaction of the incident light with the lattice vibrations in the crystal, i.e. the phonon modes. The phonon modes in crystals shift to lower frequencies with increasing temperature. The empirical form of this shift was first described by Cui *et al.*[77] and takes the form

$$\omega(T) = \omega_0 - \frac{A}{e^{{}^{hc\omega_0 B}/_{k_B T}} - 1} \tag{2.23}$$

where $\omega(T)$ is the phonon frequency as a function of temperature T, and h, k_B, and c are Planck's constant, Boltzmann's constant and the speed of light respectively. The phonon frequency at 0 K (ω_0), A, and B are fitting parameters which are material- and phonon-dependent (see Section 3.4). Figure 2.13 shows the Cui formula curves fitted to calibration data for an A_1 mode and two E_2 modes (from GaN and AlGaN layers of different Al-fractions in the same device). This device structure will be discussed further in Chapter 4.

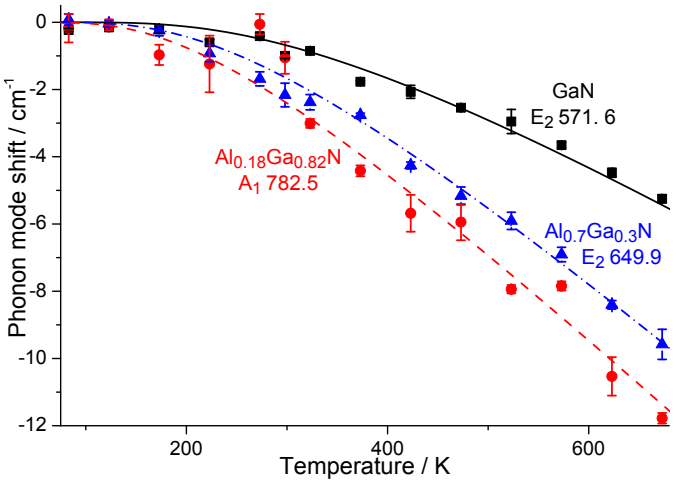


Figure 2.13: Cui formula fits for some of the phonon modes probed in the AlGaN/GaN/AlGaN DHFET discussed in Chapter 4, expressed as shifts from ω_0 which is given for each curve.

2.3 Spatial resolution in optical spectroscopy

This section discusses the theoretical limitations to the spatial resolution of the optical spectroscopy techniques used in this thesis. There are two main limiting factors which must be considered: the diffraction limit and refraction at the sample surface (leading to spherical aberration). The effects of refraction will be extended to consider some effects of polarisation and birefringence.

2.3.1 The diffraction limit

The diffraction-limited size of the focussed spot from a laser is determined by the numerical aperture of the focussing lens (i.e. the microscope objective in a micro-Raman system) and the wavelength of the light, as well as the medium in which the focus is formed. The numerical aperture, NA, is an important parameter of an optical system and is defined as

$$NA = n \sin\theta \approx n^{D}/_{2f} \tag{2.24}$$

for a refractive index n and a beam half-angle θ. The widely-used approximation applies to a lens of focal length f and aperture D, as illustrated in Figure 2.14. The following treatment assumes that the Raman system, consisting of the laser, spectrometer and microscope, can be treated as equivalent to a confocal microscope system. A confocal microscope passes the excitation light through a pinhole which is imaged onto (i.e. confocal with) the sample, and passes the return light (whether that is Raman-scattered,

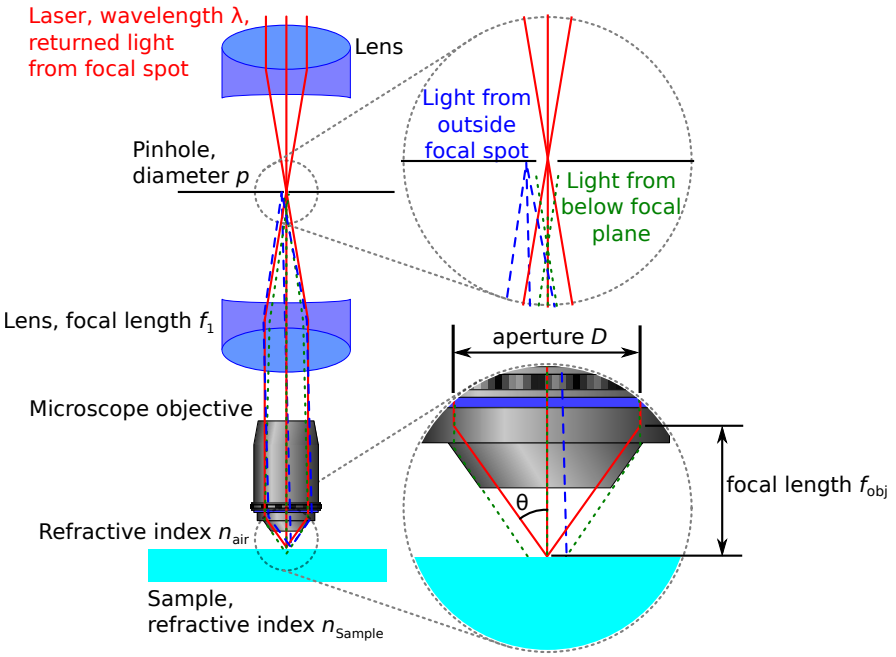


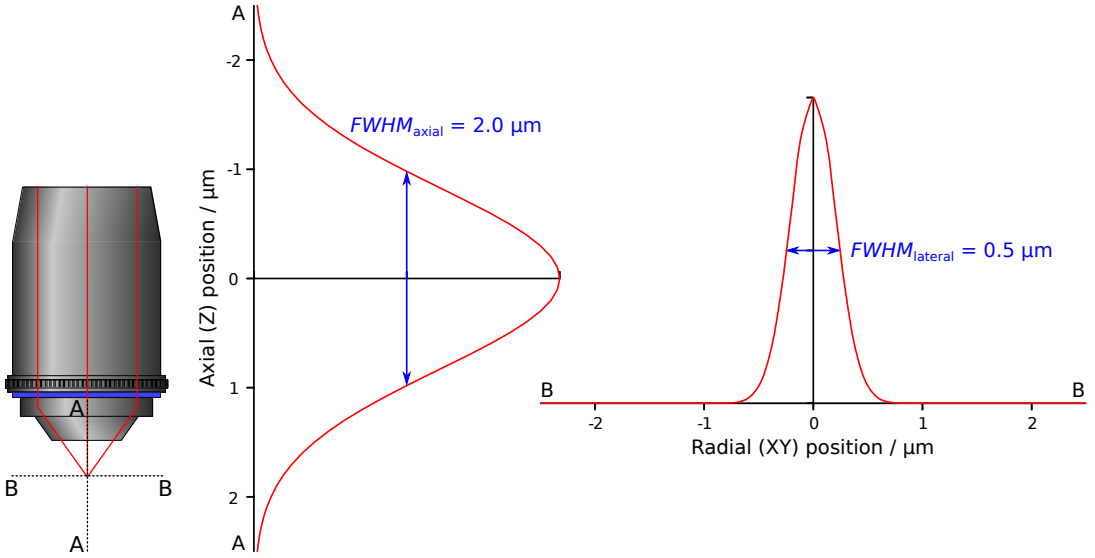
Figure 2.14: The basic principles of confocal microscopy. The laser beam (red solid lines) is focussed at the sample. Light excited by the laser returns along the same path. Light generated outside the laser focus laterally (blue dashed lines) or axially (green dotted lines) does not focus in the pinhole aperture and is therefore rejected. The upper inset shows a detail of these foci near the pinhole. The lower inset shows a detail of the objective focus.

fluorescence or simply reflected light) back through the same pinhole before measuring it. The confocal pinhole rejects light generated away from the focal plane of the objective, as this light does not form a focus at the pinhole and therefore very little of this light passes through the pinhole. It also rejects light generated within the focal plane but outside the focal spot, as while this light is refocussed at the plane of the pinhole, the focus is formed off-axis and therefore the light does not pass through the pinhole. These two effects are illustrated in Figure 2.14.

The spot size has two important dimensions, the lateral and the axial, in both of which the laser intensity distribution may be regarded as Gaussian as illustrated in Figure 2.15. It is convenient to consider the intensity full width at half maximum ($FWHM$) as being the size of the laser spot, though the width where the intensity falls to $^1/_{e^2}$ is also sometimes used. The lateral full width at half maximum is typically comparable to λ, the laser wavelength,[a] while the axial dimension is considerably larger, as illustrated in Figure 2.15 for typical parameters. They are given by[78]

$$FWHM_{\text{lateral, excitation}} = 0.51\frac{\lambda}{NA_{\text{eff}}} \tag{2.25}$$

[a]The lateral full width at half maximum given by Equation 2.25 must be modified for a focus below an interface to take into account the reduced effective numerical aperture NA_{eff}.



(a) (b)

Figure 2.15: (a) The focus in air of a microscope objective. (b) Gaussian intensity cross sections through the focal spot of a typical objective along the Z axis (along line AA) and in the XY plane (along line BB).

$$FWHM_{\text{axial}} = \sqrt{\left(\frac{0.88\lambda}{n - \sqrt{n^2 - NA^2}}\right)^2 + \left(\frac{\sqrt{2}nPH}{NA}\right)^2} \tag{2.26}$$

where n is the refractive index of the material in which the laser is focussed, NA is the numerical aperture of the objective and PH is the sample-space diameter of the pinhole in a confocal microscope system, i.e. the physical size p of the pinhole magnified by the ratio of the focal lengths of the objective f_{obj} and the lens which is focussed on the pinhole f_1. These dimensions are illustrated in Figure 2.14.

It is often desirable to limit the range of depths to which the system is sensitive, known as the confocality, for example to reduce background from layers other than those that are being measured, or to isolate the Raman spectrum from a depth slice of interest. This means limiting the axial size $FWHM_{\text{axial}}$ of the laser spot and of the volume from which the spectrometer can collect, with reference to Equation 2.26.

The importance of this is illustrated by using typical values for the systems used in this work, which will be discussed in more detail in Chapter 5. Under these typical conditions including an objective with a high numerical aperture to minimise $FWHM_{\text{axial}}$, the axial spot size in air is 2.0 µm, while in GaN it is 3.6 µm, which is nearly twice the thickness of a typical GaN buffer layer.

2.3.2 Refraction at the sample surface

The diffraction limit is only one aspect of the axial spot size formed when a laser beam is focussed below the surface of a sample. The refraction of the rays forming the laser beam when they encounter the surface of a material, such as at an air-GaN interface, depends on the angle of incidence, and therefore on the position of the ray within the beam, as illustrated in Figure 2.16. A full treatment of this spherical aberration due to refraction is given by Everall in Reference 79, but the main conclusions are summarised in the following and then extended. An ideal objective lens working in air (refractive index $n_{\text{air}} = 1$) is assumed.

After refraction at a planar sample surface, rays from different radii within the incident laser beam are focussed onto different planes. This is shown schematically in Figure 2.16 and plotted in Figure 2.17. To consider the light intensity as a function of depth inside the material, it is necessary to consider the distribution of light intensity across the cross-section of the beam, $I(m)$. For a typical laser this may be assumed to be Gaussian and so is given by

$$I(m) = I_0 \mathrm{e}^{-2m^2/\phi^2} \tag{2.27}$$

where $m = {}^{r}/_{r_{max}}$ is the normalised radius r of a ray within the beam, with $m = 0$ for a ray normal to the surface, and $m = 1$ for a marginal ray as illustrated in Figure 2.16, I_0 is the maximum intensity in the beam (the on-axis intensity), and ϕ is a fill factor which is equal to one if the objective is optimally filled i.e. if the ${}^{1}/_{\mathrm{e}^2}$ beam diameter is equal to the objective back aperture diameter $d = 2r_{max}$. The intensity profile along the axis of the beam inside the sample is therefore obtained from the parametric equation

$$\begin{aligned} I'(m) &= mI(m) \\ z_m &= \Delta \left[m^2 \frac{NA^2 \left(n^2 - 1\right)}{1 - NA^2} + n^2 \right]^{1/2} \end{aligned} \tag{2.28}$$

where $I'(m)$ is the intensity due to light from normalised radius m, z_m is the depth inside the GaN to which a ray passing through the objective aperture at m is focussed and Δ is the depth to which the laser would focus inside the sample, if the sample had refractive index 1, as shown in Figure 2.16.

It is no longer possible to define a focal *plane*; instead a focal volume must be considered. This volume will have a "centre of gravity" or centroid, the z-coordinate of which is given

by

$$\text{c.o.g.} = \frac{\int_0^1 m z_m I(m)\,\mathrm{d}m}{\int_0^1 m I(m)\,\mathrm{d}m} \tag{2.29}$$

The total laser intensity reaching the beam axis above this depth will equal that reaching the axis below it. Thus in a sense it may be considered the nearest single answer to the question of where the beam focusses, or the weighted average focus position.

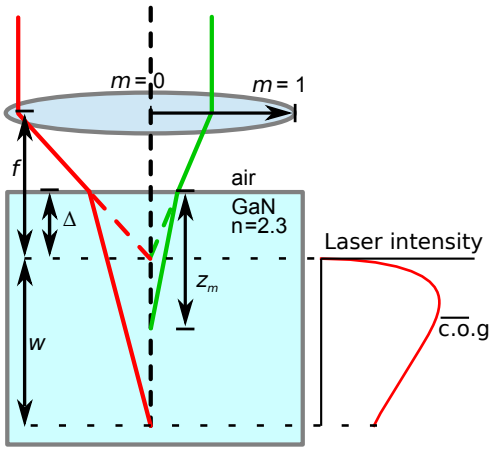


Figure 2.16: Schematic of rays focussed from air into GaN, refracted at the air-GaN interface. The objective lens focal length is f, Δ is the distance from the sample surface to the focal point, if the sample had refractive index 1, m is the normalised radial position of a ray within the objective aperture, and z_m is the depth to which a ray passing through the objective at normalised radius m is focussed. The laser intensity distribution (from Equation 2.28), spot size w (from Equation 2.32), and approximate position of the spot's centre of gravity (Equation 2.29) within the spot are illustrated. The effect of diffraction is ignored.

When applying this result to a multilayer sample such as a AlGaN/GaN HEMT a simplifying assumption can be made, namely that the refractive index of the substrate (such as SiC) is the same as that of the GaN. This is not strictly true as SiC has a refractive index of 2.55 while the refractive index of GaN is 2.39 (for polarisations perpendicular to the c-axis of the crystal in both cases), however when compared to the refractive index contrast at the air-GaN interface the refraction at the GaN-SiC interface is insignificant (also the range of angles incident on the GaN-SiC interface is much smaller than at the air-GaN interface, as a result of the refraction at the air-GaN interface, further reducing the effect at the GaN-substrate interface). Sapphire, which is also transparent, is no longer a common substrate for AlGaN/GaN HEMTs but unlike SiC it has a lower refractive index than GaN (1.76) and also a greater refractive index contrast with GaN than SiC does, so this approximation should be used with caution if applied to devices on sapphire. However the approximation would be appropriate for diamond substrates (refractive index 2.42), subject to consideration of any interlayers.

Figure 2.17 shows the distribution of laser intensity with depth z for a variety of nominal penetration depths Δ. For a high numerical aperture objective as may be chosen to minimise the diffraction-limited spot size (Figure 2.17), if the focus is more than a few microns into the GaN or substrate the length of the spot due to refraction exceeds that due to diffraction ($FWHM_{\text{axial}}$), and if data, for example Raman spectra, are acquired through the entire thickness of the wafer, as is possible with SiC wafers, the axial spot size can be many tens of microns. It may be seen in Figure 2.17 that at $m = 1$ the intensity does not fall smoothly to zero. This comes from the non-zero laser intensity at the edge of the objective's aperture, a significant departure from zero intensity considering the assumption made above that the $^1/_{e^2}$ beam width coincided with the aperture diameter. The discontinuity in the function serves as a reminder that the effects of diffraction cannot be ignored for a complete analysis.

The "width" $w_{\text{ref}} = z_1 - z_0$ of the laser spot may be defined as the distance between the planes to which rays from $m = 0$ and $m = 1$ are focussed, as found from Equation 2.28. This gives

$$z_0 = \Delta n \tag{2.30}$$

$$z_1 = \Delta \left[\frac{NA^2 \left(n^2 - 1\right)}{1 - NA^2} + n^2 \right]^{1/2} \tag{2.31}$$

resulting in a width due to refraction of

$$w_{\text{ref}} = \Delta \left(\left[\frac{NA^2 \left(n^2 - 1\right)}{1 - NA^2} + n^2 \right]^{1/2} - n \right) \tag{2.32}$$

As an example, with $\Delta = 2$ µm, as might be naively selected to measure at the GaN-SiC interface considering typical 1.9 µm GaN epilayer grown on a SiC substrate, the focus is centred at 5.7 µm with a width w_{ref} of 2.1 µm.

2.3.3 Combining the effects of diffraction and refraction on the axial spot size

To further improve on existing thermography techniques it is necessary to include not just the effect of refraction, but that of diffraction, in any calculation of the laser intensity

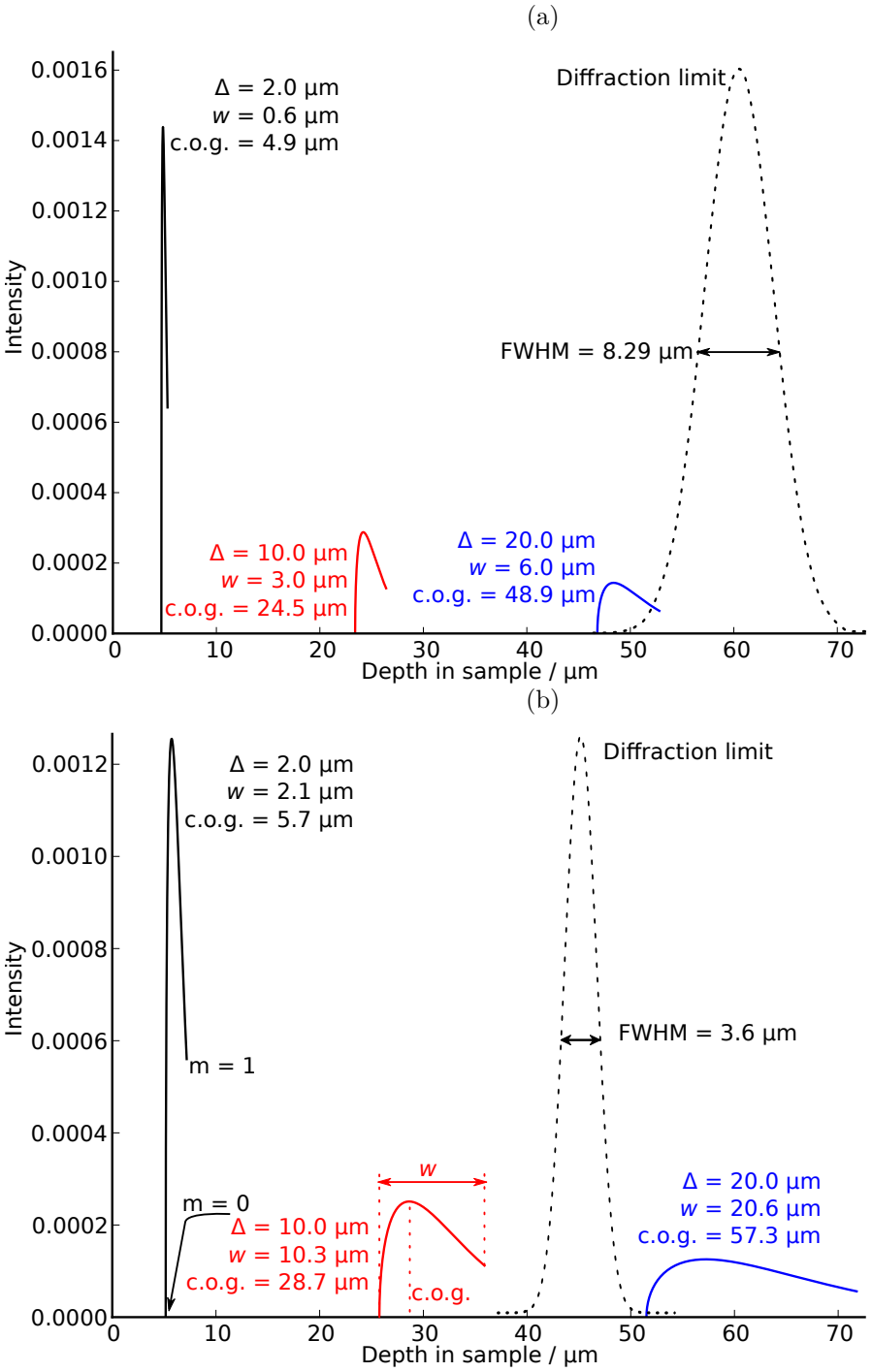


Figure 2.17: Laser intensity distribution due only to refraction for various values of Δ, the distance from the sample surface to the focal plane if the sample had refractive index $n = 1$. Curves were calculated for an objective numerical aperture of (a) 0.5; (b) 0.75, and a refractive index $n_{\mathrm{GaN}} = 2.393$. The widths w are found from Equation 2.32 and the "centre of gravity" from Equation 2.29. The curves in each plot are normalised to the same integrated intensity. Also indicated, by dotted lines, are the corresponding diffraction limits (from Equation 2.26, i.e. ignoring refraction), which once below the surface are unchanged, so are plotted at arbitrary x-positions for clarity. Defining the true distribution of laser intensity in the spot will require both diffraction and refraction to be taken into account, here their magnitudes and forms may be compared.

distribution in the sample. This will then enable calculation of the contributions from different depths to the measured weighted average temperature. A simplifying assumption is made in the approach presented here: that the patterns due to refraction and diffraction may be regarded as independent distributions. The resulting overall axial intensity profile may be calculated from a convolution of the refractive and diffractive effects It is possible to combine Equations 2.26 and 2.32 to give a reasonable approximation to the combined width of the intensity distribution due to refraction and diffraction, by treating the refraction as described by Everall as a further term to be added in quadrature to the existing two parts of the diffraction limit. However it is first necessary to define a width equivalent to the full width at half maximum for the axial intensity distribution due to refraction. From Figure 2.17, it can be seen that the concept of full width at half maximum is meaningless for spherical aberration as the intensity distribution does not go continuously to zero, and is not symmetric. As can be seen from Figure 2.18(a) the width w is appropriately compared to the $^1/_{e^2}$ width, d_{1/e^2}, of a Gaussian, which is related to the full width at half maximum by $FWHM = \frac{\ln(2)}{\sqrt{2}} d_{1/e^2}$, rather than to the full width at half maximum directly.

Including this result as a third component to be added in quadrature to those in Equation 2.26 gives a depth resolution (axial spot size) below the surface of

$$w_{\mathrm{ax}} = \sqrt{\left(\frac{0.88\lambda}{n - \sqrt{n^2 - NA^2}}\right)^2 + \left(\frac{\sqrt{2}nPH}{NA}\right)^2 + \frac{\ln 2}{2}\Delta^2\left(\left[\frac{NA^2\left(n^2-1\right)}{1-NA^2} + n^2\right]^{1/2} - n\right)^2} \tag{2.33}$$

This is compared to the convolution of the refractive and diffractive effects on the axial spot size for $\Delta = 2$ µm and 10 µm in Figure 2.18(b). For small values of Δ the approximation is acceptable, but as the effect of refraction begins to dominate that of diffraction it becomes less appropriate. At which values of Δ this occurs depends on the numerical aperture, as a high numerical aperture increases w_{ref} while decreasing $FWHM_{\mathrm{axial}}$). It is also often necessary (for experiments in which the depth varies) to consider the spot centre shift, for which Equation 2.29 may still be used.

The effect of this spherical aberration can be greatly reduced when working at a fixed (or nearly fixed[a]) depth by using an objective with variable coverslip correction, which can be adjusted to produce an equivalent and opposite aberration by translating internal elements of the objective.[80] Although this is designed for a glass coverslip with

[a] A depth may be regarded as nearly fixed if changes in Δ are small compared to Δ.

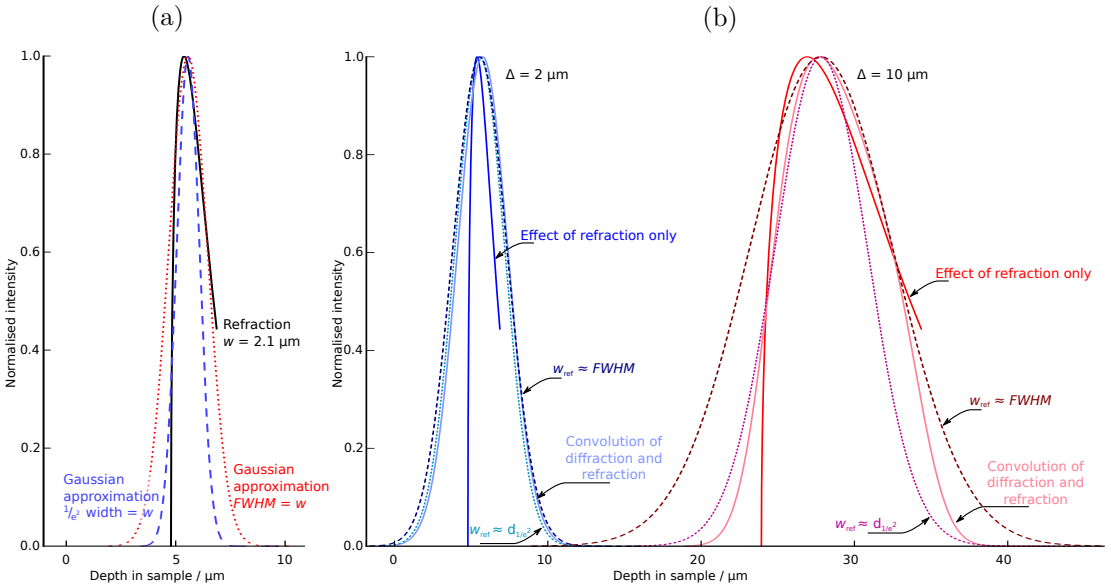


Figure 2.18: Approximating the axial spot profile as a Gaussian: (a) Gaussian approximation to the effect of refraction at $\Delta = 2$ µm, comparing the approximations $FWHM \approx w_{\mathrm{ref}}$ (red dotted curve) and $d_{1/e^2} \approx w_{\mathrm{ref}}$ (blue dashed curve); (b) Gaussian approximation to the convolution (light solid curves) of the effects of refraction (dark solid curve) and diffraction at $\Delta = 2$ µm and 10 µm. Dotted curves show $d_{1/e^2} \approx w_{\mathrm{ref}}$ and dashed curves show $FWHM \approx w_{\mathrm{ref}}$.

a refractive index n_{glass} of ~1.52, by selecting a coverslip thickness of $\frac{n_{\mathrm{sample}}}{n_{\mathrm{glass}}}\Delta$ much of the benefit of the coverslip correction is regained for fixed values of Δ.

Equations 2.27 to 2.33 take no account of Fresnel reflection at the sample surface, which for a high index material such as GaN combined with a high numerical aperture objective will lead to a significant difference in reflection for $m = 0$ and $m = 1$ and so modify $I'(m)$, the distribution of light given in Equation 2.28. The effect of Fresnel reflection is polarisation dependent, even on an isotropic sample, as the Fresnel reflection depends on polarisation, however it does not change the width w defined above, merely shifting the centre of gravity of the laser spot slightly towards the surface for p-polarised light and away from the surface for s-polarised light compared to the results given above neglecting Fresnel reflection. Fresnel reflection can therefore be neglected in calculating the axial spot size using the approximation in Equation 2.33, however it should be taken into account if using the convolution for a more exact calculation, as it affects the distribution of light along the length of the laser spot.

2.3.4 Extending spherical aberration to birefringent materials

When applied to transparent device-relevant materials such as GaN, SiC, and sapphire, a further complication is introduced: these materials are birefringent, i.e. their refractive

indices are anisotropic. In the case of GaN, the optic axis is parallel to the *c*-axis, i.e. normal to the surface for the typically-grown HEMTs studied in this thesis. The ordinary index of refraction n_{o}, which applies to light polarised perpendicular to the *c*-axis is lower than the extraordinary index of refraction n_{e} for light polarised parallel to the *c*-axis.[81] The refractive index for a given polarisation and wavelength may be calculated from the Sellmeier dispersion relation,[82] which may be given in a number of forms, one of which, for a single absorption frequency, is[81]

$$n^2(\lambda) = A^2 + \frac{B^2}{\lambda^2 - C^2} \tag{2.34}$$

where λ is the wavelength for which the index of refraction is to be found, and A, B, and C are fitting parameters, which will be different for n_{o} and n_{e}. Table 2.2 gives the calculated refractive indices for GaN.

Equations 2.27 to 2.33 ignore the effects of birefringence, which unlike Fresnel reflection does affect the axial spot size. Figure 2.19 shows how incident linear polarisation experiences different refractive indices when it encounters a birefringent sample. Light polarised perpendicular to the plane of incidence always experiences n_{o}, while the refractive index experienced by light polarised parallel to the plane of incidence, or at intermediate angles, depends on both the polarisation angle with respect to the plane of incidence and the angle between the propagation direction and the *c*-axis.

For polarisation in the plane of incidence (p-polarisation) it is therefore necessary to modify Equation 2.28 to take into account the angle-dependent refractive index found from the index ellipsoid by[82,83]

$$\frac{1}{n(\theta)^2} = \frac{\cos^2(\theta)}{n_o^2} + \frac{sin^2(\theta)}{n_e^2} \tag{2.35}$$

Table 2.2: Parameters[81] for the Sellmeier relation for GaN for $\mathbf{E} \perp c$ and $\mathbf{E} \parallel c$, along with the values for n_{o}, n_{e} and the birefringence $n_{\mathrm{o}} - n_{\mathrm{e}}$ at the wavelengths of interest in this thesis.

	n_{o}	n_{e}	$n_{\mathrm{o}} - n_{\mathrm{e}}$
A	2.25	2.30	
B (nm)	330.6	368.2	
C (nm)	259.7	268.9	
$\lambda = 488$ nm	2.328	2.393	−0.065
$\lambda = 532$ nm	2.318	2.381	−0.063

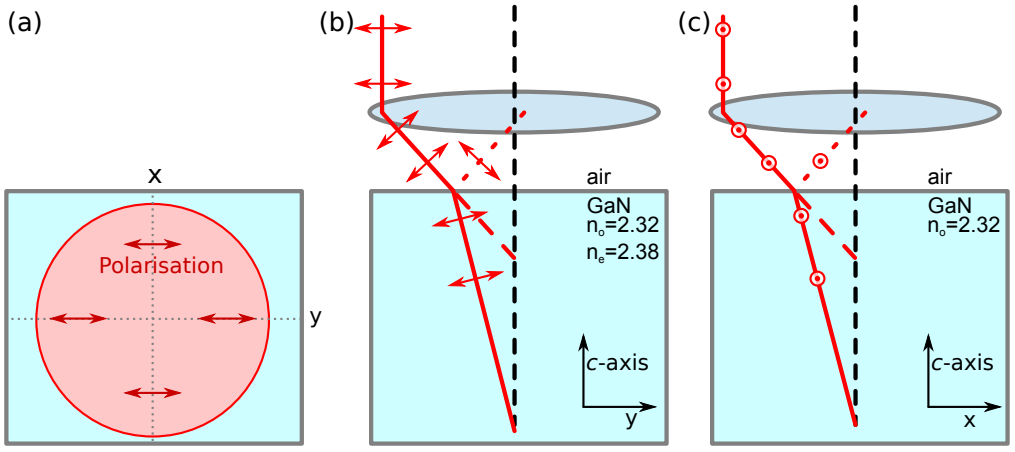


Figure 2.19: The effect on incident linear polarisation of Fresnel reflection and a birefringent sample. (a) Top view along the incident beam, showing the polarisation with respect to the x- and y-sample axes; (b) View along the x-axis, light polarised parallel to the y-axis is p-polarised, i.e. polarised parallel to the plane of incidence. As different rays within the beam are refracted to different angles with respect to the c-axis, they are subjected to different refractive indices; (c) view along the y-axis, polarisation is still along the y-axis but is now s-polarised – perpendicular to the plane of incidence. These rays all experience the same refractive index n_o.

where θ is the angle between the propagation direction of a ray and the optical axis of the crystal (the c-axis for GaN and SiC). However for s-polarisation it is only necessary to replace n with n_o in Equation 2.28 as the polarisation remains perpendicular to the c-axis. The two polarisations parallel and perpendicular to the plane of incidence represent the extreme cases, as may be seen be considering a plane intermediate to those illustrated in Figure 2.19 (b) and (c). The axial spot size w_{ref} from Equation 2.32 is therefore greater for p-polarised light than for s-polarised light, and greater still for a beam comprising both s- and p-polarisations, such as a linearly polarised beam converging towards a focus. It is simplest to consider the s- and p-polarisation states separately as representing the limits of the problem for large Δ, if a correction is made in the objective for the isotropic spherical aberration. If Δ is small and no spherical aberration correction is used, the additional effect on the spot size due to the birefringent sample is small compared to the dimension found from Equation 2.33, as may be seen in Figure 2.20.

2.3.5 Alternative polarisations

It might be thought that circular polarisation would provide a solution to the adverse effects of birefringence discussed above, however circularly polarised light still experiences different refractive indices depending on the angle of incidence. It is therefore necessary to consider a system such that all rays behave as shown in Figure 2.19(c), In such a

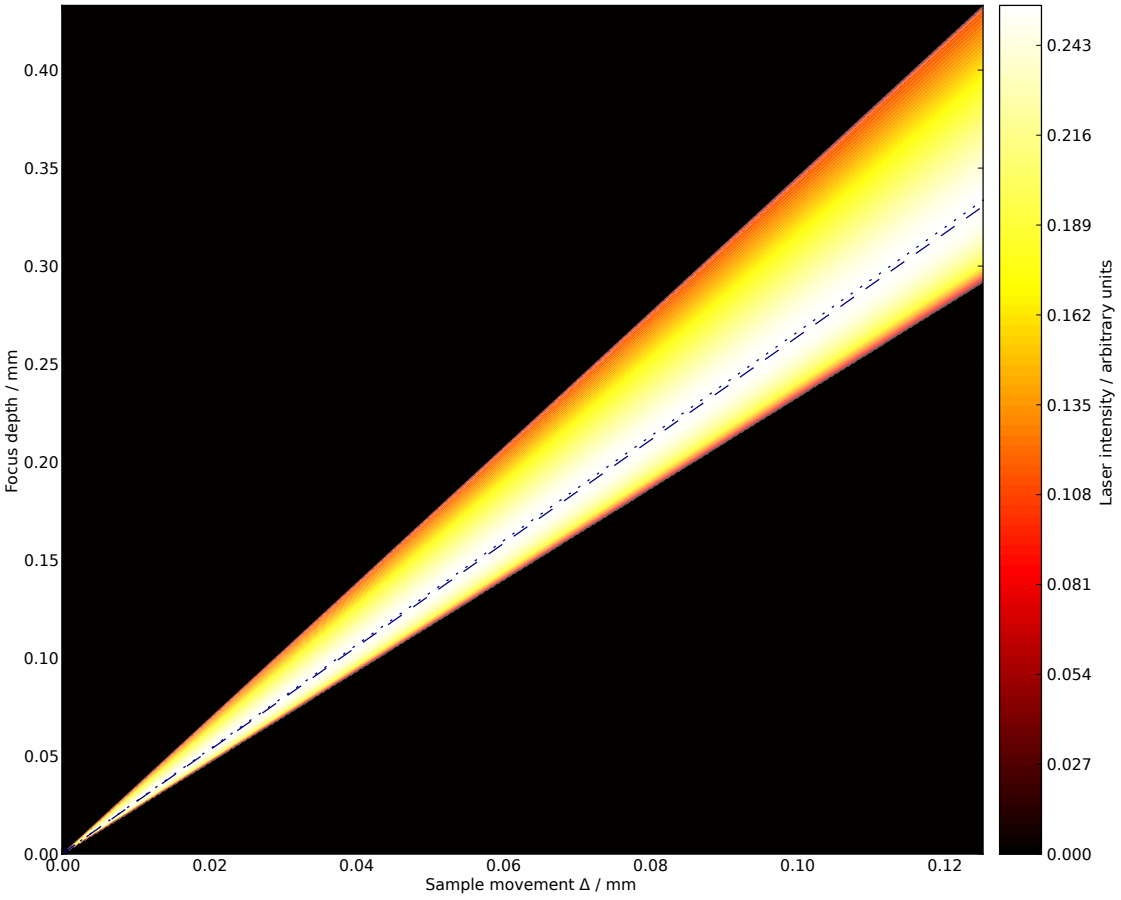


Figure 2.20: Laser intensity distribution as a function of Δ and depth in the sample for linearly polarised light, taking into account Fresnel reflection and birefringence. The dark blue lines show the maxima for s- (dashed) and p- (dotted) polarisations.

system the polarisation vector of any ray must always be orthogonal to the beam axis, even as the beam converges to a focus. This equates to azimuthal polarisation.

Several ways of achieving azimuthal polarisation have been developed, such as a focussing along the optic axis of calcite – a highly birefringent crystal – followed by a spatial filter[84,85] (which will be discussed in Chapter 5), and spatially-varying waveplates.[86,87] A development of the last technique is to use a femtosecond laser to create nanostructures in silica.[88] This component is commercially available under the name *s-wave*, and its use will be discussed in Chapter 5. These approaches may also be used to achieve radial polarisation, which is of interest for its small lateral spot size compared to more conventional polarisations, as well as for generating z-polarisation, i.e. polarisation along the beam axis at a focus, as will be shown in Section 5.1.5. In general the same optical systems may be used to generate both radial and azimuthal polarisations, though there are additional options for generating radial polarisation.[89]

2.3.6 Limitations of current Raman thermography techniques

At present, while Raman thermography may be used to give a high-resolution map of temperature in the XY plane of a device, it is less useful for measuring temperature as a function of depth within typical device epilayers, as has been seen in Sections 2.3.1 to 2.3.4. In fact, as the temperature may not be considered constant with depth in a GaN buffer, Raman thermography must be interpreted with care – the spot size calculations presented here allow the depth-weighting of temperature measurements to be calculated. In Chapter 4 one method of obtaining a higher temperature resolution is explored, which requires some specific sample features. In Chapter 5, however, the theory presented in this section is applied to a more general approach to collecting temperature information from a range of depths, with the goals of a more direct measurement of certain device parameters, and of placing more precise constraints on device simulations.

2.4 Photoluminescence spectroscopy and thermography

Photoluminescence may also be used to gain information, such as temperature, from semiconductors. Above-bandgap excitation is required because photoluminescence relies on an electronic excitation within the sample, which means sufficient photon energy to excite electrons into the conduction band is needed, as can be seen in Figure 2.10. A laser above the bandgap of the material(s) being studied, but usually below the bandgap of any intervening layers, is used. The photoluminescence emission rate is governed by Einstein's A and B coefficients (Equation 2.18) in a similar way to the rate of Raman scattering, giving a total emission rate for transitions from state n to state m of[70]

$$R_{nm} = A_{nm} + B_{nm}\rho_e(\nu) \tag{2.36}$$

where $\rho_e(\nu)$ is the photon energy density. Equation 2.36 is more helpfully rewritten in terms of the photon occupation number N_p as

$$R_{nm} = A_{nm}(1 + N_\mathrm{p}) \tag{2.37}$$

The second term, proportional to N_p, determines the rate of stimulated emission, and may be neglected outside a laser cavity, i.e. assuming that the photon density $\rho_e(\nu)/\hbar\nu$ is small.

Considering conduction (c) and valence (v) bands instead of levels n and m, the emission rate for recombination form the conduction band to the valence band is given by

$$R_{\mathrm{cv}} = A_{\mathrm{cv}} f_{\mathrm{c}} (1 - f_{\mathrm{v}}) \tag{2.38}$$

where f_{c} and f_{v} are the electronic occupancies in the conduction and valence bands, therefore $(1 - f_{\mathrm{v}})$ is the probability of a valence band state being empty. The inverse process of absorption causing an excitation from the valence to the conduction band, denoted by P_{vc} is given by

$$P_{\mathrm{vc}}(\nu)\rho(\nu) = R_{\mathrm{vc}}(\nu) \tag{2.39}$$

at thermal equilibrium (the van Roosbroeck–Shockley relation), and is related to the absorption coefficient α by $P_{\mathrm{vc}} = {}^{\alpha c}/_{n}$, where c is the speed of light in vacuum and n is the refractive index of the material. The photon density $\rho(\nu)$ is dependent on temperature according to the Planck distribution, which allows the emission probability inside the sample, neglecting re-absorption effects, to be related to the absorption coefficient by

$$R_{\mathrm{vc}}(\nu) = \frac{8\pi\alpha(\nu)\nu^2 n^2}{c^2 \left[e^{\left({}^{h\nu}/_{k_B T}\right)} - 1 \right]} \tag{2.40}$$

This is a useful relation, though it only strictly holds for intrinsic semiconductors, as dopants have various effects on the presence and occupancy of states in or near the band gap. The factor $\mathrm{e}^{\left({}^{h\nu}/_{k_{\mathrm{B}}T}\right)}$ in Equation 2.40 leads to an exponential decrease in carrier population with increasing energy. This enhances the emission from low-energy states, making photoluminescence (and related techniques, as will be shown in Section 2.5) sensitive to low-lying energy levels, which may include radiative recombination states within the gap. At thermal equilibrium, the derivation above holds for any pair of states, not just the band-to-band process described, however at thermal equilibrium the total emission is insignificant unless the temperature is very high, so photoluminescence uses an excitation source to produce non-equilibrium conditions. Generally speaking, although a non-thermal distribution is excited by the incident light, the relaxation into a thermal distribution is fast compared to the electron-hole recombination time. The process of photoluminescence, as was illustrated in Figure 2.10, thus consists of three steps: *excitation* of electron-hole pairs by an energy source; *thermalisation* or relaxation to a thermal distribution; and radiative *recombination*. The same spectral features may also be excited by other means, such as cathodoluminescence (CL) excited by an electron

beam, or electroluminescence excited by electric fields and current within the material being studied. Electroluminescence will be discussed further in Section 2.5. The characteristic spectral features in luminescence experiments do not change in frequency with the excitation source, unlike the features studied in Raman spectroscopy.

2.4.1 Band-to-band transitions and photoluminescence thermography

In a direct bandgap semiconductor, for example GaN, because there is no change in momentum between the valence band maximum and the conduction band minimum, recombination of thermalised excitons is highly probable, at a frequency equivalent to the bandgap energy. Such materials are important for applications such as LEDs. The rate of radiative recombination of free electrons and holes is governed by[70]

$$R_{\mathrm{rad}} = {}^{n_{\mathrm{e}} n_{\mathrm{h}}}/_{\bar{\tau}_{\mathrm{r}}} \tag{2.41}$$

where n_{e} is the free electron concentration, n_{h} is the free hole concentration, and $\bar{\tau}_{\mathrm{r}}$ is the average radiative recombination time, given the thermal distribution of electron and hole energies.

In photoluminescence, electrons and holes are excited in pairs, i.e. in equal numbers, and in large numbers such that $n_{\mathrm{e}} = n_{\mathrm{h}} \gg n_{\mathrm{i}} = p_{\mathrm{i}}$, where n_{i} and p_{i} are the intrinsic electron and hole concentrations respectively. The radiative recombination rate ${}^{1}/_{\tau_{\mathrm{rad}}}$ is then given by

$${}^{1}/_{\tau_{\mathrm{rad}}} = {}^{n}/_{\bar{\tau}_{\mathrm{r}}} \tag{2.42}$$

where $n = n_{\mathrm{e}} = n_{\mathrm{h}}$

As nonradiative recombination is also possible, in which the energy of the electron-hole pair is dissipated by phonons as heat, the total decay rate is given by

$${}^{1}/_{\tau_{\mathrm{tot}}} = {}^{1}/_{\tau_{\mathrm{rad}}} + {}^{1}/_{\tau_{\mathrm{non\text{-}rad}}} \tag{2.43}$$

where τ_{tot} and $\tau_{\mathrm{non\text{-}rad}}$ are the total and non-radiative recombination rates respectively. The band-to-band emission spectrum of a direct-bandgap semiconductor with bandgap E_{g} is governed by the joint density of states

$$D_{\mathrm{j}} \propto \sqrt{E - E_{\mathrm{g}}} \tag{2.44}$$

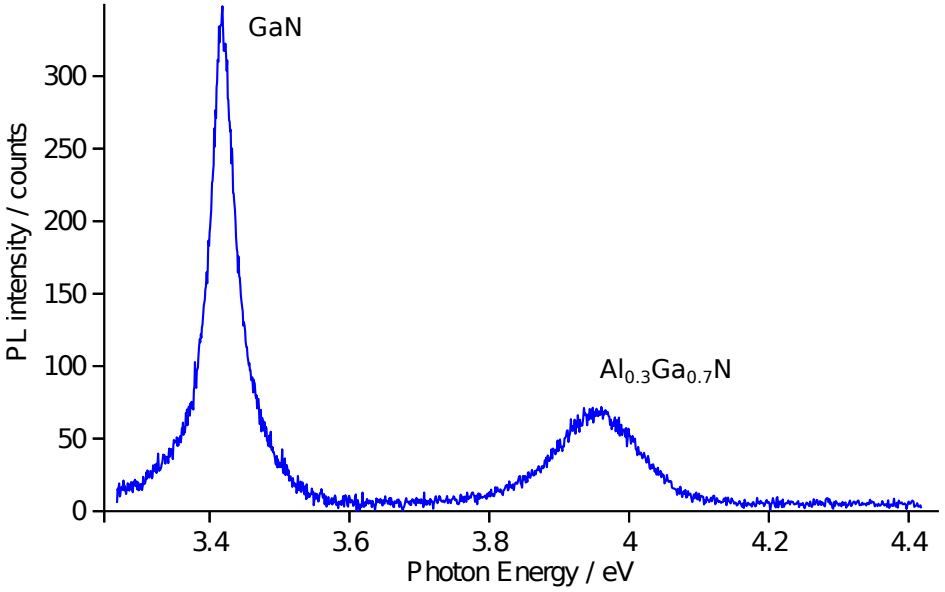


Figure 2.21: Photoluminescence spectrum measured with a 244 nm laser located at the University of Bath on an AlGaN/GaN HEMT, showing peaks at the band edges of GaN and AlGaN. This spectrum was taken as part of the photoluminescence experiment discussed in Chapter 6.

and the occupancy distribution functions of electrons and holes. The electron occupancy f_e, in the quasi-equilibrium case is given by the Boltzmann distribution and may be written as

$$f_\mathrm{e} \propto \mathrm{e}^{\left(-E/k_\mathrm{B}T\right)}. \tag{2.45}$$

The hole distribution f_h may be defined in the same way.

Substituting Equations 2.44 and 2.45 into Equation 2.38 gives a photoluminescence spectral line shape of

$$I_\mathrm{PL}(\hbar\omega) \propto \begin{cases} \sqrt{\hbar\omega - E_\mathrm{g}}\,\mathrm{e}^{-\hbar\omega/k_\mathrm{B}T} & \text{for } \hbar\omega > E_\mathrm{g} \\ 0 & \text{otherwise} \end{cases} \tag{2.46}$$

for emitted photon energy $\hbar\omega$. Figure 2.21 shows a typical photoluminescence spectrum, of a GaN buffer and AlGaN barrier. The characteristic peaks, at the band-edge positions expected from Figure 2.4 are broadened by the finite instrument resolution.

Photoluminescence thermography probes the change in bandgap with temperature, a quantity that is independent of the exciting laser used. Several formulae have been proposed to fit the change in bandgap E_g with temperature T. The empirical equation proposed by Varshni in 1967[90] is still widely used for convenience, and was used for the

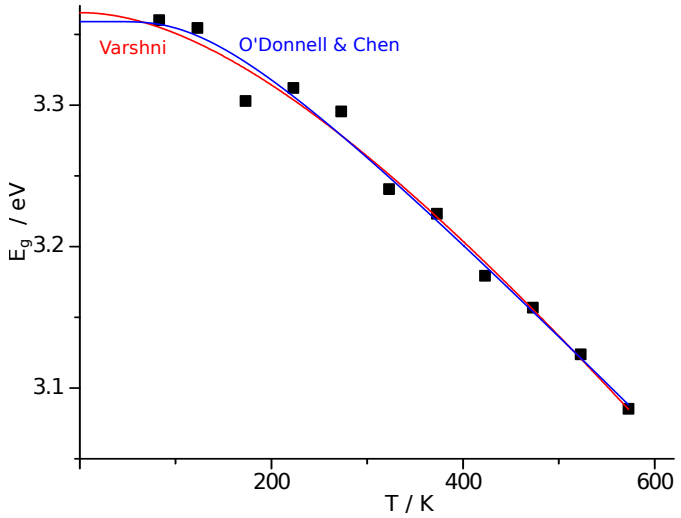


Figure 2.22: Measured (black squares) GaN band gap as a function of temperature. The solid lines shows the Varshni equation fit (red) and the fit using the more physically-relevant formula proposed by O'Donnell and Chen[91] (blue).

analysis in this work

$$E_g = E_0 - \frac{\alpha T^2}{(T+\beta)} \tag{2.47}$$

where E_0 is the bandgap at 0 K, and α and β are material-dependent constants, all of which which must be determined experimentally (see Section 3.4). The band edge peak position of the GaN channel layer of the devices discussed in Chapter 4 is shown in Figure 2.22 along with the Varshni equation fit to the data. Varshni's equation does however lack the connection to the underlying physics of more recent models,[91] the deviation being particularly important at cryogenic temperatures, which are not uncommon in photoluminescence studies of device-relevant materials. This more appropriate formula models the bandgap as

$$E_g(T) = E_0 - S\overline{\hbar\omega}\left[\coth\left(\overline{\hbar\omega}/2k_BT\right) - 1\right] \tag{2.48}$$

where S is a coupling constant, and $\overline{\hbar\omega}$ is an average phonon energy; k_B, as usual, is Boltzmann's constant. At the elevated temperatures used in this work it is clear from Figure 2.22 that Varshni's formula closely follows O'Donnell and Chen's and will not introduce a significant additional error in the interpretation of photoluminescence thermography results.

2.4.2 Photoluminescence for defect analysis

When photoluminescence is excited by a continuous wave laser, an equilibrium condition may be assumed, so that the total decay rate $^1/_{\tau_{\text{tot}}}$ is unchanged if the laser power is unchanged. If the band-to-band recombination photoluminescence signal level, determined by $^1/_{\tau_{\text{rad}}}$, changes, then from Equation 2.43, a corresponding change in $^1/_{\tau_{\text{non-rad}}}$, the non-radiative transition rate, must occur, suggesting the generation of non-radiative recombination centres. This will be discussed in Chapter 6 in the context of off-state stress and its effects on device degradation. Deep Ultraviolet photoluminescence was used to ascertain whether damage due to off-state stress (observed via electroluminescence) was confined to the buffer or affected the barrier, by comparing the magnitude of the photoluminescence peaks from the two layers.

2.5 Electroluminescence

Electroluminescence (EL) is light emitted due to an electric current or field. *Recombination* electroluminescence, well known from LEDs, but also possible in other electronic devices such as HEMTs, takes place over a narrow range of wavelengths, as electrons from the conduction band recombine with holes in the valence band. The energy is released as a photon. The recombination may be band-to-band, at the bandgap energy as shown in Figure 2.23, or it may involve dopants or defect states, for example due to donor-acceptor pairs leading to electroluminescence at other energies.

Recombination electroluminescence has many parallels with photoluminescence (and indeed with cathodoluminescence), the difference having more to do with the means of excitation than the resulting emission. In contrast to the situation in GaAs, the wide bandgap of GaN and AlGaN makes band-to-band recombination highly unlikely in HEMTs, due to the thermal occupancy factor in Equation 2.40. It has however been used to demonstrate the occurrence of impact ionisation in AlGaN/GaN HEMTs with InGaN back-barriers, where the recombination takes place at the lower InGaN bandgap energy.[93] Donor-acceptor pair transitions may take place at lower energies, and are more likely to be observed in electroluminescence spectra.

When a semiconductor contains both donors and acceptors it is said to be *compensated*, because at equilibrium some of the electrons from the donors will be captured by the

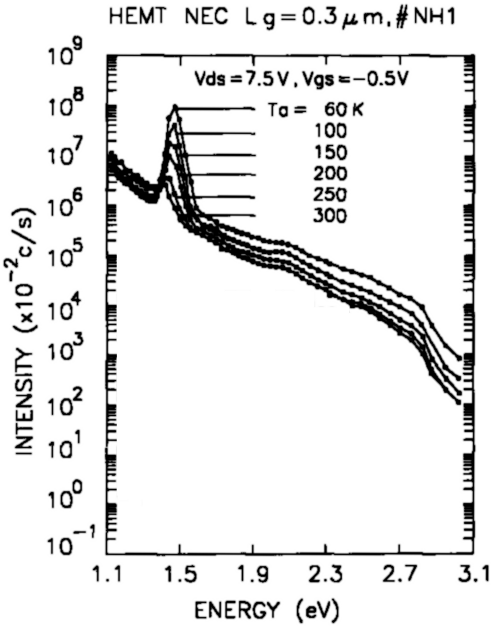


Figure 2.23: Electroluminescence from a GaAs HEMT showing the recombination peak at the bandgap superimposed on a continuum of hot-electron emission for a range of device temperatures at $V_{\mathrm{DS}} = 7.5$ V and $V_{\mathrm{GS}} = 0.5$ V. As the device temperature is decreased the hot-electron temperature and intensity increase, while the band-edge recombination peak shifts as expected (Equations 2.47 and 2.48. Figure taken from Reference 92.

acceptors.[70] A compensated semiconductor contains both ionised donors (D^+) and acceptors (A^-). Optically or thermally excited electrons (in the conduction band) or holes (in the valence band) may then be trapped at the D^+ and A^- states producing neutral (D^0 and A^0) centres. If equilibrium is restored by an electron on a neutral donor recombining with a hole on a neutral acceptor a photon may be emitted, by the reaction

$$\mathrm{D}^0 + \mathrm{A}^0 \rightarrow \hbar\omega + \mathrm{D}^+ + \mathrm{A}^- \tag{2.49}$$

The photon energy depends not just on the donor and acceptor levels (E_{D} and E_{A} respectively) but on the energy of the Coulomb interaction between the ionised donors and acceptors, which lowers the energy in the final state of Equation 2.49 by $^{e^2}/_{\epsilon R}$ where e is the charge on an electron, ϵ is the dielectric constant and R is the separation between the donor and acceptor. As the final energy is lowered, the photon energy is increased, and is given by[70]

$$\hbar\omega = E_{\mathrm{g}} - E_{\mathrm{A}} - E_{\mathrm{D}} + {}^{e^2}/_{\epsilon R} \tag{2.50}$$

Electroluminescence emission spectra may be used to investigate the behaviour of dopants in operating devices[94] by comparison with spectral features observed through other luminescence techniques.[95] At room temperature, donor-acceptor pair features are often

broad, though at cryogenic temperatures discrete lines may be observed in some samples. Even these broad features may however be attributed to particular states in some cases.

2.5.1 Hot electron electroluminescence

The high field present in some semiconductor devices accelerates electrons to the point that their kinetic energy, expressed in thermal terms, greatly exceeds that of the lattice. An "electron temperature" can therefore be defined, which is higher than the lattice temperature, i.e. the conventional temperature of the semiconductor material. This is particularly true in HEMTs as a result of the decreased scattering (higher electron mobility) in the 2DEG compared to devices based on bulk semiconductor properties. This leads to electroluminescence when the electrons are eventually scattered and lose energy in the form of photons. The use of hot-electron electroluminescence in investigating degradation in both on- and off-state devices will be explored in Chapter 6.

The usual methods for calculating drift velocity are not valid when the carrier energy gain is rapid enough that the carriers and the lattice are not in thermal equilibrium.[70] Hot carriers are however in thermal equilibrium amongst themselves, and a carrier temperature may be defined, described by a Fermi-Dirac distribution. This defines the equilibrium probability f_k of a state at energy E being occupied by a carrier at a carrier temperature T in the absence of an applied field as

$$f_k = \frac{1}{\mathrm{e}^{(E-E_\mathrm{F})/k_\mathrm{B}T} + 1} \tag{2.51}$$

where as before E_F is the Fermi energy level and k_B is Boltzmann's constant.

Hot electrons behave in a qualitatively different way to those in equilibrium with the lattice. The drift velocity for hot electrons saturates, so the behaviour of the semiconductor is no longer Ohmic. In GaN, negative differential resistance is also observed. These two phenomena are shown in Figure 2.24.

For the electron drift velocity to saturate as the field increases requires an energy loss mechanism that becomes important at electron kinetic energies. This may be explained by inelastic scattering involving optical phonons.[70] There is still some debate about the precise mechanisms involved in the emission of light from hot electrons in GaN-based devices, though bremsstrahlung involving phonon interactions (necessary for momentum

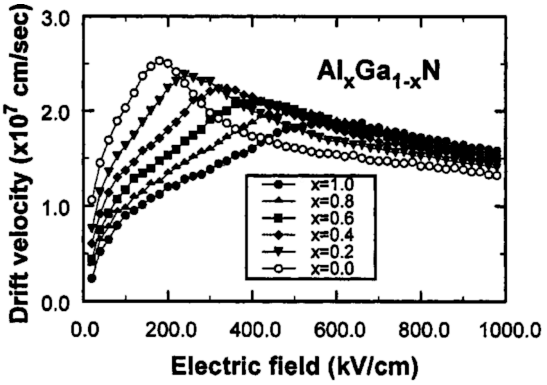


Figure 2.24: The drift velocity of electrons in GaN and AlGaN as a function of applied field, showing negative differential resistance (in which the drift velocity decreases with increasing field) and saturation, a seen in the current-voltage curves in Figure 2.7. Figure taken from Reference 96.

conservation) would allow both intra- and inter-valley scattering[97] (i.e. scattering within and between valleys in the conduction band, which are shown in Figure 2.3) as has been reported for many years in other semiconductors such as GaAs[98] and ZnSe.[99]

As hot-electron electroluminescence is dependent on a high electric field to accelerate the electrons, it may be used experimentally to probe the regions of high field[100] within a device under a wide range of operating conditions. As would be naively expected, the hot-electron electroluminescence intensity increases with drain bias V_{DS}, but its dependence on gate bias V_{GS} is nonlinear, as the electroluminescence intensity is a function of both field and current. As the gate-source voltage V_{GS} is increased from the pinch-off voltage the intensity will first increase (as current starts to flow) and then decrease (as the peak field near the gate reduces). This leads to a bell-shaped curve in a plot of electroluminescence intensity against V_{GS}.[101,102]

In an off-state device where the electric field is high but there is ideally no current, regions of localised emission – "hotspots" – are therefore of interest as locations of localised current flow and therefore potentially device degradation.[103,104] This will be discussed further in Chapter 6, as previous work has demonstrated the existence of off-state electroluminescence spots, and looked into the conditions required for them to form, but an understanding of the underlying mechanisms has proved elusive. The work discussed in Chapter 6 therefore sets out first to precisely locate and then to understand the processes involved in off-state degradation, by studying the electroluminescence emitted from sites of gate leakage and applying photoluminescence to them for the first time.

Chapter 3

Experimental details

3.1 Optical systems

The majority of the experiments discussed in this thesis made use of optical spectroscopy or imaging. In this section the instruments used for spectroscopy (electroluminescence, photoluminescence and Raman) and for imaging (white light and electroluminescence) will be described, along with details specific to the techniques themselves and certain experimental configurations.

3.1.1 Spectrometer and microscope systems

The theory behind both Raman and photoluminescence spectroscopy, as well as electroluminescence, was set out in Sections 2.2 to 2.5 but it is necessary, before discussing experiments and their results in detail, to describe some of the instruments, their configuration and use. The optical measurements discussed in this thesis were carried out using Raman microscope systems made by Renishaw plc., from the *RM 1000* and *inVia* series. As the inVia is based on the RM 1000, the systems are very similar, and the same description will suffice for both. The main difference is increased automation in the inVia, though the software – *WiRE 2.0* – for the inVia system is more capable than for the RM 1000 – *WiRE 1.3*. Figures 3.1 and 3.2 show the main components of the Raman microscope system, as configured for a typical thermography experiment.

The laser light enters the system from behind the spectrometer, and is passed through a beam expander to enable it to fill the back aperture of the objective optimally. The light from the laser must in most cases be passed through a filter to remove unwanted spectroscopic lines due to atomic transitions in the lasing medium, known as plasma lines, as well as other unwanted spectral features. Although these transitions lead to

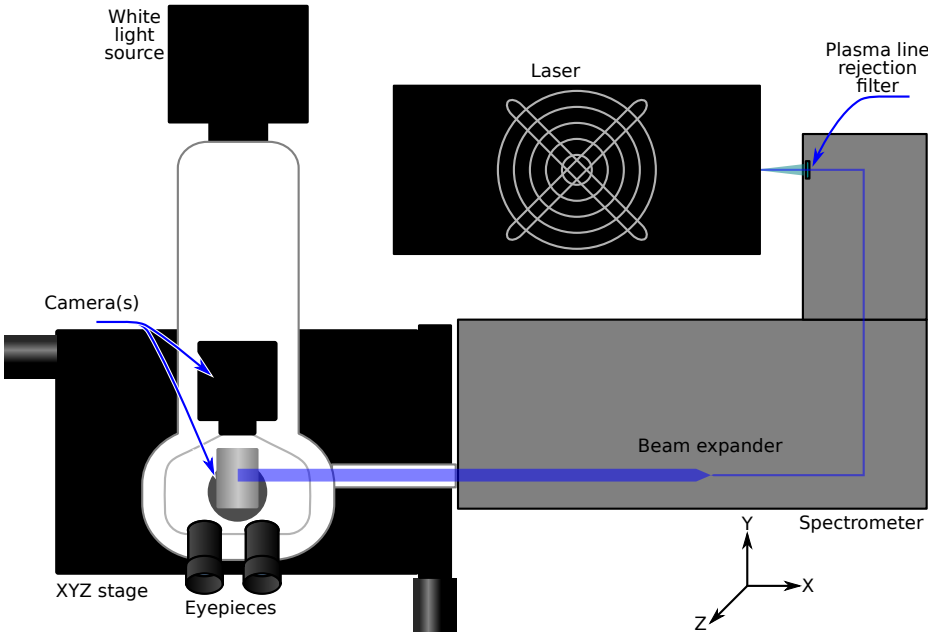


Figure 3.1: Plan view of the Renishaw RM series Raman spectrometers. The axes shown are those used for describing sample movement.

divergent light and are inherently weak compared to the laser line, they can still appear stronger than the Raman peaks from many materials.

After the plasma line rejection filter the laser beam is reflected by the laser rejection filters ("Rayleigh filters"), which may be long-pass ("edge") or stop-band ("notch") filters, both of which have high reflectivity at the laser wavelength and high transmissivity for Raman-scattered light. Notch filters are required for Stokes-anti-Stokes measurements, while edge filters are more durable and give better signal-to-noise ratio in some experiments. Edge filters are also preferred for photoluminescence as they have a flatter transmission profile over the wider spectral range typically used for photoluminescence. For electroluminescence spectroscopy the filters are removed, as no laser is used, and the filters would block some wavelengths which might be of interest. Regardless of the type of filter used, the filters must transmit no more than 10^{-12} to 10^{-9} of the laser light, and block as little as possible of the Raman-scattered light. To achieve this level of blocking two filters are combined in a set.

After the laser rejection filters the excitation light passes through the microscope before being focussed on the sample by the objective. It is possible, by the use of switching mirrors and beamsplitters, to configure the microscope for white-light imaging at the eyepieces or cameras. All systems are fitted with cameras and eyepieces to allow the sample to be viewed and aligned. These also allow the laser spot to be viewed when

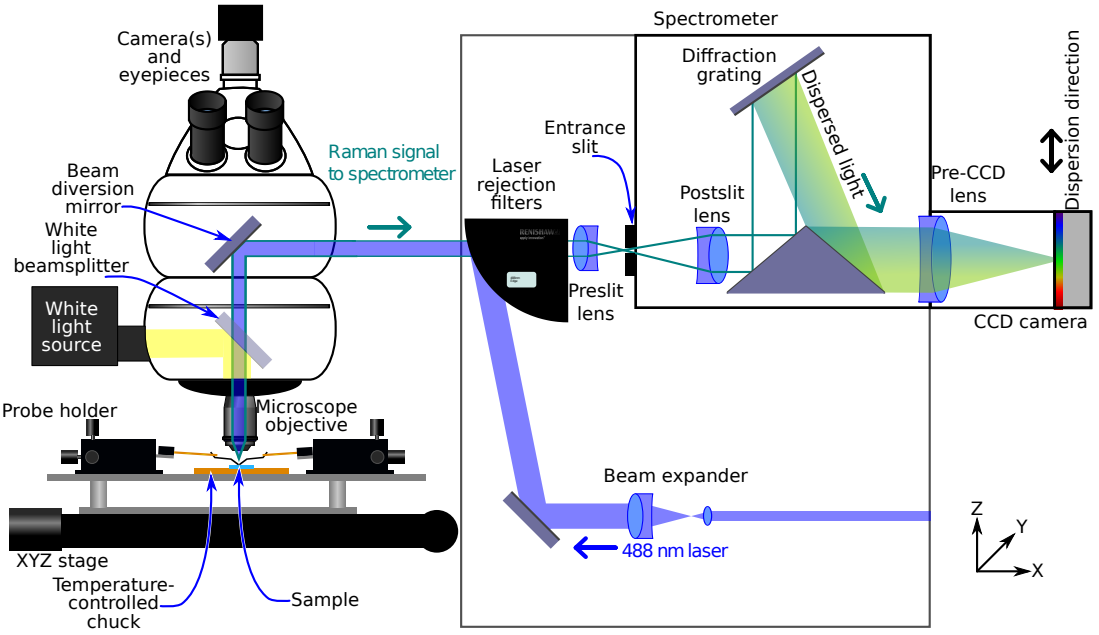


Figure 3.2: The general layout of the Renishaw RM series Raman spectrometers, configured for Raman thermography at 488 nm. The axes shown are those used for describing sample movement. The spectrometer entrance slit is parallel to the Y axis, i.e. perpendicular to the plane of the figure, while the CCD spectral dimension is vertical. For photoluminescence the configuration is similar: the laser wavelength changes and the beam expander is not present. For electroluminescence the laser is not used and the filters are removed.

used with visible or near infrared lasers, but not with ultraviolet lasers. Some systems have had high-sensitivity cameras added for electroluminescence imaging (Figure 6.4(b) for example) which may be selected instead of the eyepieces and default camera. These are not shown in Figure 3.2, but are located behind the eyepieces.

In a spectroscopy experiment the objective collects the Raman-scattered light, photoluminescence or electroluminescence and directs it towards the spectrometer. Rayleigh scattered light at the laser wavelength is blocked by the filters. The light is focussed onto the adjustable entrance slit and recollimated before being dispersed by the diffraction grating and imaged on a CCD camera. Multiple diffraction gratings may be fitted to a given system. The choice of which grating to use depends on the desired spectral resolution (determining the grating groove density, or number of lines per millimetre), and the operating wavelength. The effects of the entrance slit width and CCD image area of the depth resolution of the system are discussed in Section 5.1.

The CCD image area is shown schematically in Figure 3.3. It consists of a spectral dimension, along which the light is dispersed by the diffraction grating, and a much smaller dimension which holds purely spatial information. Thus a “column” on the CCD

consists of light all at the same wavelength[a] from all points imaged through the slit, while a "row" consists of different wavelengths from the same point. The signal for a user-selected number of pixels in each column is summed to produce a data point in the spectrum.

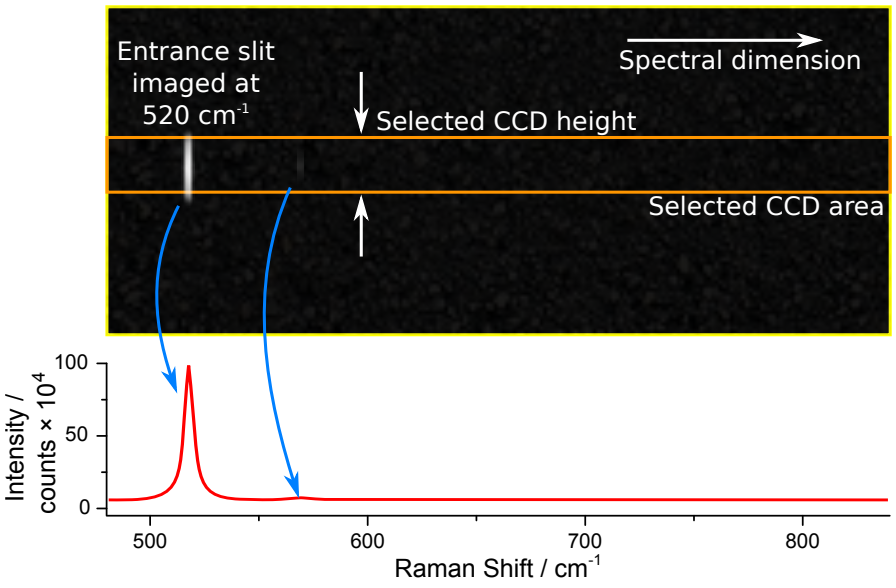


Figure 3.3: The CCD image area and its relation to the Raman spectrum.

The CCD image sensor itself is a silicon semiconductor device consisting of an array of photosensitive capacitors, each of which forms one pixel.[105] These capacitors are formed by a layer of insulating material separating a p-type Si layer from a metal or polysilicon gate. Unless the CCD sensor is *back-illuminated*, i.e. unless the Si thickness has been reduced until it is thin enough to transmit light to the active region, the gate must be semi-transparent, in which case the incident light arrives from the gate side. A bias is applied to the gates of pixels which are to be used, causing a depletion region to form. This permits selective shuttering, the importance of which will become clear in Section 5.1. Under illumination, electron-hole pairs are generated when incident photons are absorbed, and the electrons are drawn towards the gate, charging the capacitor. At the end of the exposure, the array must be read out. This is done by transferring the charge on the pixels to an output register, using the CCD as a shift register – its original purpose when first designed in 1970. In the detector used in the Renishaw systems described here, the output register consists of one potential well per CCD column, where a column corresponds to light at a narrow range of wavelengths. This allows the signal from a number of pixels within a column to be summed within the CCD, i.e. as analogue data, which reduces shot noise (noise due to statistical variations in the number of

[a] Within the constraints of the spectral resolution.

incident photons), while discarding spatial information orthogonal to the system slit. The pixels of the output register are read out one-by-one, by passing the charge on them to a charge amplifier. Once again at this stage, multiple (output register) pixels may be summed before being read out, a feature known as *binning*. This again improves signal-to-noise ratio, this time at the expense of spectral resolution. In the design of a Raman spectrometer considerable effort and expense is employed to maximise the spectral resolution. The maximum resolution may not always be required (for example in electroluminescence spectroscopy) and binning provides a way to trade off resolution against the signal-to-noise ratio that can be obtained in a given time. The output of the charge amplifier is a voltage, which is then converted to a digital signal. In these steps, further noise is inevitably introduced. In addition, any electron-hole pairs generated for any reason other than photo-generation in the active layer are a source of signal background and noise, known as dark-current. This is minimised by cooling the CCD; in the case of the Renishaw CCD used here a stack of Peltier (thermo-electric) heat pumps achieves -70 °C.

As well as analogue signal integration on-chip, multiple exposures, or *acquisitions* were used in some experiments. Each acquisition is read out independently after it is complete, and the data are added pixel-by-pixel in software. Although this increases the noise per data point due to readout and signal quantisation compared to adding on-chip, it enables a higher dynamic range, important when strong and weak peaks are present in the same spectrum, as the strong peaks would *saturate* before adequate signal-to-noise could be obtained on the weak peaks. Saturation in the sense of CCDs means that pixels at the highest point of the peak are unable to register more photons, as their wells are full. This leads to a flat-topped appearance in the spectrum and makes it impossible to determine with any precision the peak height or the central peak frequency, a variable of critical importance in Raman thermography.

The CCD's ability to shift charge is further exploited in another mode of operation available on the Renishaw Raman systems – continuous extended scanning. In this mode, referred to by the manufacturer as *synchroscan* the charge on the CCD is moved in the spectral dimension, synchronously with the movement of the diffraction grating, and while light is being collected on the CCD. This allows a wider spectral range to be measured than would otherwise be possible for a single grating, and is free of the artefacts that would appear in a so-called *step-and-repeat* scan in which the grating is moved in

discrete steps between exposures. Unlike using a lower groove density diffraction grating, extended scanning allows the full spectral resolution of the system to be maintained. An exposure using extended scanning is time-consuming, taking a total time t_{total} given approximately by

$$t_{\text{total}} \sim \left(\lceil \frac{N_{\text{readout}}}{N_{\text{CCD}}} + 1 \rceil t_{\text{exp}} + t_{\text{readout}} \right) N_{\text{acq}} + t_{\text{move}} \left(N_{\text{acq}} - 1 \right) \tag{3.1}$$

where N_{readout} is the total number of pixels to be read from the CCD, determining the spectral range, N_{CCD} is the number of pixels in the spectral dimension of the CCD, t_{exp}, t_{readout} and N_{acq} are the exposure time, readout time and the number of acquisitions respectively, and t_{move} is the time taken for the grating to move from the position where it finishes one acquisition to where it starts the next. Not only does N_{readout} depend on the spectral range to be measured, but so does t_{move}. The readout time, $t_{\text{readout}} \sim 1$ s and only varies slightly with the CCD area. To put this into context, for a single acquisition ($N_{\text{acq}} = 1$) for a typical electroluminescence experiment in which $N_{\text{readout}} \sim 9000$ (without binning) and a 60 s exposure time, $t_{\text{total}} \sim 17$ min. This would have to be repeated several times in a realistic experiment to allowing averaging for noise-reduction. On the modern inVia systems, $t_{\text{move}} \sim 1$ min, but on the older RM models, $t_{\text{move}} \sim 10$ min, thus an experiment involving extended scans can easily take over an hour to give a single spectrum, once averaging is taken into account. As extended scanning is mainly used for experiments in which a wide spectral range is required, such as photoluminescence and electroluminescence, it is often combined with binning, thus spectral resolution is discarded anyway, as high resolution is rarely required in such experiments, and the additional signal-to-noise ratio may be helpful. Binning also adds a useful level of flexibility compared to selecting from a limited range of gratings.

A further reason for multiple acquisitions is the removal of *cosmic ray* features. Despite the name many of these are actually due to gamma rays given off by atomic decay in materials such as concrete.[106] These appear randomly in spectra as sharp, often strong, peaks, and are therefore more likely if a larger CCD image area is used and for longer acquisition times. They are problematic if automated peak detection methods are to be used, as they can be stronger than the spectral peaks from the sample, which may cause an analysis routine to find a cosmic ray peak in preference to real data. In many cases the best way to deal with cosmic ray peaks is simply to retake the measurement, particularly if the likelihood of a cosmic ray peak appearing is low per acquisition. In

more automated experiments, or more time-sensitive ones, this is not always possible, and the peaks must be dealt with as outliers. They may be manually removed from particular spectra, alternatively the WiRE software includes a routine to remove such features by taking two extra acquisitions which do not contribute to the final output. The exact method used is not specified, but is likely to be based on discarding extreme values on a pixel-by-pixel basis. Other methods, such as acquiring several spectra and basing the analysis on the median[a] signal value for each pixel (i.e. for each wavelength) can give better signal to noise in the same amount of time as the Renishaw software at the expense of some loss of automation of the process. This semi-manual approach is particularly useful for electroluminescence spectroscopy, in which a wide image area and long acquisition time may be used, leading to many cosmic rays, and the signal is weak enough that every acquisition should count towards the final data. Any method used to remove cosmic ray spikes must work pixel-by-pixel, relying on multiple spectra. This is because an approach based on spectral filtering, taking into account multiple pixels within one spectrum, is likely to have an effect on real spectral features. Smoothing is useful in some situations, but as a method for removing cosmic ray spikes it generally introduces artefacts without fully removing the spikes.

3.1.2 Optical experiments on operating devices

For thermography (Raman or photoluminescence) or electroluminescence the sample, whether a packaged device or a wafer-piece, was usually mounted on a temperature-controlled chuck. An operating device may dissipate several watts of heat in the on state, where it may be kept for some time. The chuck must therefore be able to remove this heat while maintaining a stable temperature. Thermo-electric cooling modules (Peltier devices) were used in many thermography experiments, combined with water-cooling for high electrical powers. For some low-power experiments passive heatsinking was sufficient.

For most experiments the device was contacted using probe tips designed for DC use (needle probes) or for RF use, which have a screen making contact to the source of the device. Two RF probes or three DC probes were required for connecting to the three device contacts on a HEMT. Figure 3.2 shows DC probes with one omitted for clarity. As will be seen in Chapter 5, the need for probe tips in thermography experiments places

[a] A similar approach is to use the mean value, with outliers detected and excluded.

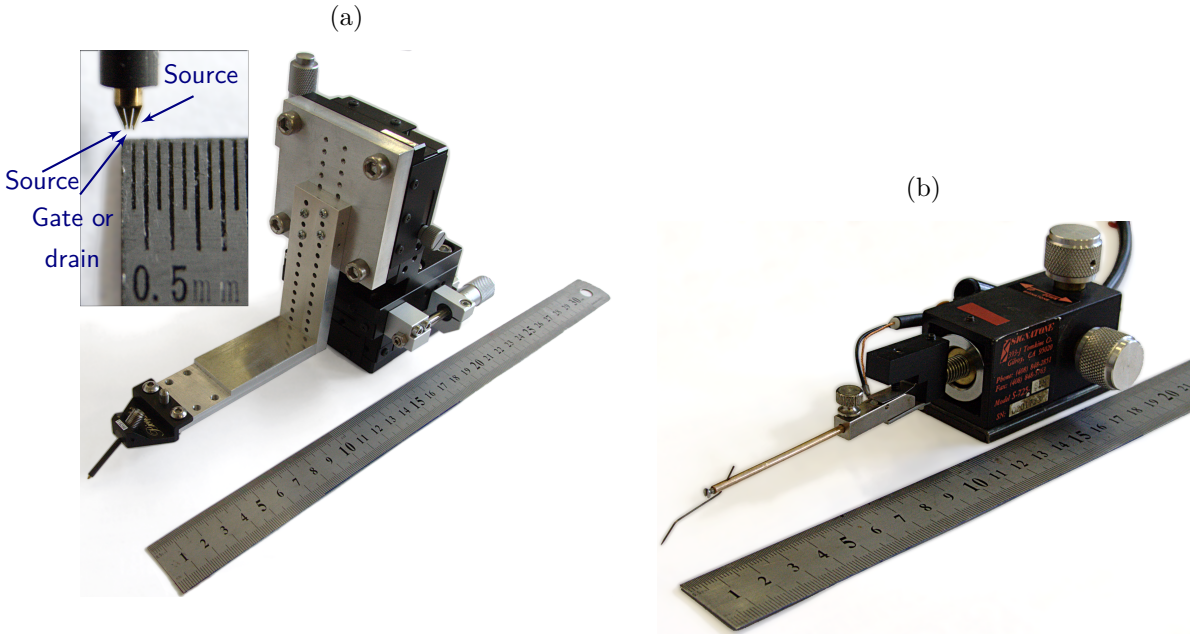


Figure 3.4: Probes used to make electrical contact to devices: (a) RF probe (the inset shows a detail of the probe tip); (b) DC (needle) probe. The ruler scale is in cm in the main figures.

some constraints on the sampling optics. RF probe tips are larger than DC probe tips, and more fragile. The simpler DC probe tips may be formed to occupy less space than the manufacturer intended, so were therefore generally used in this work. The probe tips appear as dark areas on top of the device contacts in white light images of devices, as will be shown in (for example) Figure 6.4. The probes provided sufficient pressure to hold some devices firmly against the sample chuck, when combined with a viscous heat transfer compound. In other cases a vacuum chuck was used to provide an additional holding force.

Not all devices measured were contacted using probes – some devices were *wire-bonded*. The thermosonic wire-bonding process works by applying ultrasound vibrations combined with pressure to a gold wire in contact with a heated workpiece (such as the contact pad of a device). The temperatures used are much lower than the melting points of the materials used in the devices, or the (usually gold) wire. This process may be carried out manually or (semi-)automatically. Manual wire-bonding was used for certain measurements in which probe tips could not be applied to the device. Some devices were provided in a packaged form. In such devices wire-bonding had already been used to connect the device to contacts leading outside the package, suitable for connection to an electronic circuit. For optical measurements open packages were used, or the lids of the packages omitted.

3.1.3 Mapping

The XYZ motorised stage (Figure 3.2) provided submicron sample positioning in three dimensions on inVia systems. The older RM 1000 systems were equipped with an XY stage only, with a manual Z (focus) stage. The system software could, for both types of system, interface with both the spectrometer and the XY(Z) stage, allowing the sample to be moved precisely between spectral acquisitions. The result of this process was a *map* in one or two dimensions. These maps could consist of Raman, electroluminescence, or photoluminescence spectra. The term *"area maps"* is used to describe a map in the XY plane; *line maps* or *linescans* are one-dimensional maps along X, Y or an intermediate direction; and *depth scans* or *depth maps* are one-dimensional maps in Z. By default the finest available resolution in Z was 1.0 µm; some of the depth mapping in Chapter 5 required finer steps, which were achieved by reconfiguring the software. *Depth slices* are two-dimensional maps with a Z component and a component in the XY plane, but were not used in this work. A *time series* is analogous to a map in the time dimension. On the older RM systems, as there was no focus (Z) motor, maps with a Z component were not possible.

As each map consisted of spectra at many points, and each spectrum could take several minutes, there was a time dimension to all mapping experiments. This was not merely a matter of experimental convenience but could affect the results if for example the device under test aged during a measurement. In addition over long periods the systems were susceptible to drifting due to vibrations or thermal changes in the ambient environment. As a result, for many experiments devices under test had to be biased under conditions that would be unlikely to cause degradation over the course of the experiment, especially as in many cases it was necessary to repeat an experiment on the same device. Some spectral acquisitions took even longer (electroluminescence spectra using extended scanning and multiple acquisitions could take over an hour). These were still combined with time series measurements in some experiments, to measure the off-state electroluminescence spectrum as it developed over several days of stressing.

3.1.4 Imaging and spectroscopy from the back of the wafer

The typical measuring configuration shown in Figure 3.2 was used for measuring devices from the top (i.e. the contact side). This is conventional, both because it is simple to

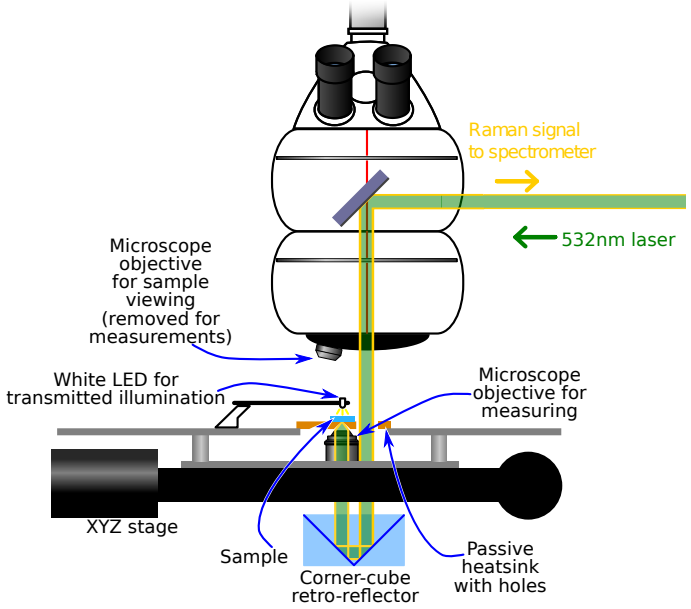


Figure 3.5: Modifications to the standard configuration shown in Figure 3.2 for measuring through the wafer using a retro-reflector. DC or RF probes may still be used, but are omitted for clarity.

contact the device under test with probe tips and because some substrates (such as Si) are opaque. With transparent substrates such as SiC and sapphire (Al_2O_3) measurements through the wafer were possible. To measure wafers from the back required a system of mirrors or a retroreflector, shown in Figure 3.5, and a heatsink with a hole positioned under the device being measured, precisely aligned to the microscope objective. Some means of viewing the device from the top in order to contact it with probe tips was also required, such as moving the stage under a low-magnification objective. Focussing through the wafer led to a considerable loss of spatial resolution, as well as a loss of signal, and such experiments were more challenging than from the top due to space and alignment constraints. This approach did, however, allow optical access under the gate and any field plates. Because there was only space for a passive heatsink, only experiments in which low electrical power was dissipated could be carried out using this configuration. Most of the work in this thesis that used the retroreflector was off-state electroluminescence, where the power dissipation is no more than a few milliwatts, in on-state experiments such as may be seen in Figure 6.3, where the power dissipation was 2.2 W, the temperature of the device would increase during a measurement. In this case on-state measurement times were minimised and recovery time was allowed between measurements, to minimise the temperature rise.

3.2 Raman thermography of GaN, AlGaN, and substrates

Some of the advantages of Raman thermography were described in Section 2.2.5, but applied to real experiments on real devices there are some additional constraints, which are considered here. It is important to select one or more suitable phonon modes for each material in the sample to be measured. If the sample is a typical HEMT or transfer length method (TLM) test structure, these material are the GaN buffer and the substrate (metals, such as device contacts, are not compatible with Raman thermography). A more complex layer structure gives rise to additional modes from the materials making up the extra layers. Considering GaN, Section 2.2.4 showed that the Raman tensors permit detection of the E_2 and A_1 phonon modes in $z(x-)\bar{z}$ configuration, and that the E_2 phonon splits into *low* and *high* branches. Experimentally the E_2 (low) peak is too weak to be of any use, leaving the A_1 and E_2 (high) modes available for thermography. The E_2 (high) mode is more sensitive than the A_1 mode to in-plane strain (including strain due to thermal expansion mismatch between the substrate and the epilayers) as it consists of in-plane vibrations. As can be seen in Figure 2.13, the A_1 mode also shifts more with temperature than the E_2 (high). For these reasons the A_1 mode is preferred for thermography. The same modes are present in AlGaN as in GaN, however they are complicated by broadening due to alloy disorder and, in the case of the E_2 mode, splitting into AlN- and GaN-like lines. This will be explored further in Chapter 4.

As GaN and related materials are piezoelectric, an electric field will induce a strain in the crystal and therefore a shift in the Raman peak corresponding to any phonon mode sensitive to that strain. Therefore in a realistic thermography experiment it is general practice to consider the change in Raman peak position from pinched off to on state,[a] as the field in a pinched off device is similar to the field in an on state device, but there is no self-heating when pinched off, making a pinched-off measurement a suitable reference. The peak shift from pinched off to the on state is then subtracted from the peak position at ambient temperature found using the Cui formula (Equation 2.23); the resulting position is then used to calculate the temperature in the on state, using the Cui formula again. When measuring on transfer length method test structures, which have no gate, it was not possible to acquire pinched-off spectra, instead the unpowered peak

[a] Recalling from figure 2.7 that in this context on-state measurements are carried out with no load, and with bias conditions intended to cause self-heating.

position was used as a reference position. It is therefore particularly important to use the A_1 phonon for thermography on transfer length methods. As the E_2 (high) phonon mode gives a much stronger signal than the A_1 (LO), some experiments that would not be possible using the A_1 mode due to insufficient signal-to-noise ratio are enabled by the use of the E_2 (high) Raman peak, for example those in Chapter 5.

Not just the buffer but the substrate of a HEMT (or transfer length method test structure) can be probed, as the common substrates also show clear phonon modes in their Raman spectra, which also shift with temperature. In the case of Si the mode at 520 cm^{-1} is used while for SiC the mode at 776 cm^{-1} is clearest – these are room temperature values. For devices on silicon, Raman spectroscopy using a visible laser means excitation above the bandgap of the substrate (see Figure 2.4), which leads to absorption of the laser light in the top of the substrate with associated localised heating. The measured substrate temperature is also therefore that of the very top of the substrate. A SiC (or sapphire) substrate is rather different: the substrate is transparent to the laser, which greatly reduces sample heating and means that the measured substrate temperature for a laser focus at the sample surface is a weighted average extending some distance in to the substrate, discussed in Section 2.3 and Chapter 5.

3.3 Photoluminescence

For photoluminescence experiments, above-bandgap excitation must be used, which in the case of thermography of GaN-based devices should also be below the bandgap of the AlGaN barrier, unless the barrier is to be probed as well, in which case the laser must be above the bandgap of this too. Two ultraviolet spectrometer systems were used for the photoluminescence discussed in this thesis, both were Renishaw RM 1000-series Raman microscopes. The spectrometer system used for photoluminescence thermography (Chapter 4) used a $\lambda = 325$ nm laser, which, as can be seen in Figure 2.4 is above the bandgap of GaN but below that of most AlGaN compositions, and thus will pass through a typical AlGaN buffer layer. When investigating the reasons for hotspots appearing in electroluminescence images of pinched off devices, as described in Chapter 6, it was necessary to investigate the AlGaN barrier as well, meaning that a deep ultraviolet laser operating at $\lambda = 244$ nm was required, located at the University of Bath. The photon energy of this laser is sufficient to interact with AlGaN compositions of all but

the highest aluminium fractions, however as the AlGaN barriers of the devices studied using this system were so thin ($\sim\lambda/10$) absorption in the barrier was incomplete, and a signal was obtained from the GaN buffer as well.

As was illustrated in Figure 2.10, photoluminescence causes an electronic excitation within the sample. When measuring operating semiconductors care must be taken to avoid these excitations having an adverse effect on the results, as they may, for example, increase the current flowing in a device, which may even lead to localised damage. Laser power densities must therefore be kept as low as practicable when measuring devices with a bias applied – as electronic excitations are localised to the laser spot it is the peak power density rather than the total deposited power that is most important.[76] A large laser spot (compared to Raman measurements) may therefore be beneficial in some cases.

The microscope objectives that must be used for photoluminescence, especially on operating devices when space for probe tips is required, give a larger laser spot than those used for Raman spectroscopy. The range of materials available to designers of ultraviolet lenses is small compared to work in the visible (mainly quartz and CaF_2, while in the visible many glass compositions are available). This restricts the performance of ultraviolet lenses, in particular the numerical aperture. In addition, the market for ultraviolet optics is tiny compared to that for visible optics, so the total range of products available is smaller. The objectives used in this work are listed in Table 3.1.

A further issue with ultraviolet photoluminescence spectroscopy is that the ultraviolet RM 1000 systems used did not have beam expanders fitted, making the working numerical aperture smaller than the specified numerical aperture of the objectives as their back apertures were not fully filled by the laser beam. The photoluminescence spot size was typically 5 µm to 7 µm, depending on the objective used, compared to submicron spots in the visible. It was also necessary to correct the microscope focus for the chromatic aberration in the objective – after focussing in white light the focus stage was adjusted by an objective-dependent correction, which had to be be empirically determined. Any errors in this offset would further increase the spot size.

When used for thermography of real devices, in which the heating is localised leading to a significant temperature gradient in the optical access region of the device, this large (compared to visible lasers) spot size could lead to a significant underestimate of the

channel temperature. This underestimate was quantified from a simulated (Section 3.5) device channel temperature, though a Raman-measured temperature could equally have been used as it has submicron spatial resolution. Spectra were recorded as a function of temperature using backplate heating as described in Section 3.4. At each point in the channel a spectrum was synthesised from a weighted sum of spectra acquired under background heating conditions. The weight attached to each spectrum was the fraction of the laser intensity incident on regions of the device with a temperature matching that spectrum. Thus for each point along a line from source to drain a synthesised spectrum was produced, from which a temperature could be found by curve fitting. The advantage of this method over a simpler approach of calculating a weighted average based purely on temperature and laser intensity (as used for Raman thermography in Chapter 5) is that the shape and intensity of the photoluminescence peak both change with temperature over the range of temperatures observed in an operating device, which could not easily be taken into account in calculating the average temperature. In other words, for a lateral dimension x perpendicular to the gate, with a laser intensity $L(x)$ (calculated from the beam profile, which was taken to be Gaussian) the signal $S'(E, x)$ at a photon energy E in the synthetic spectrum may be found from

$$S'(E,x) = \sum_x \frac{L(x)S(E,T(x))}{L(x)} \tag{3.2}$$

Table 3.1: A comparison of visible and ultraviolet objectives used for work in this thesis.

Objective	Magnification [a]	NA [b]	working distance [c] / mm	Usage
Leica N Plan L 50×/0.5[107] [d]	50	0.5	8.2	Visible (including Linkam) [e]
Leica N Plan 50×/0.75[108]	50	0.75	0.5	Visible (no space for probes)
Zeiss LD Plan-Neofluar 63×/0.75 Corr Ph2[109]	75	0.75	2.2 [f]	Visible (DC probes only)
OFR [g] LMU-15×-NUV[110] [c]	15	0.32	8.5	ultraviolet (including Linkam)
OFR LMU-40×-NUV[110]	40	0.5	1	ultraviolet (DC probes only)

[a] When used with a tube lens of 200 mm as on Leica microscopes.
[b] Numerical aperture.
[c] Working distance.
[d] Only these objectives could be used for samples inside the Linkam hot/cold cell (section 3.4).
[e] The Linkam cryostat was used for temperature calibration and low-temperature measurements; its use will be described in section 3.4.
[f] With no coverslip; reduces to 1.2 mm for a 1.5 mm glass coverslip.
[g] Optics for Research (now part of Thorlabs, inc.).

where $S(E,T)$ is the signal at photon energy E from a recorded spectrum at temperature T. The position-dependent temperature $T(x)$ was taken from a simulation and the apparent temperature $T'(x)$ was found by applying the Varshni equation (Equation 2.47) to S'. Figure 3.6 shows the result of applying this approach to the device studied in Chapter 4. In the simulation, the self-heating caused by operating the device at $V_{\mathrm{DS}} = 20$ V and $V_{\mathrm{GS}} = 1.6$ V was modelled as a fixed heater under the gate overhang. The steps visible in the simulated-photoluminescence curve are artefacts of the finite temperature steps used when collecting the original photoluminescence spectra $S(E,T)$. This demonstrates that in devices where the temperature changes significantly on length scales less than the photoluminescence laser spot size (such as HEMTs) photoluminescence thermography will always underestimate the peak temperature. While the exact amount will depend on device geometry, the underestimate is likely to be significant, and will be greater than any underestimate from Raman thermography for realistic device geometries. For the conditions used to generate Figure 3.6, the underestimate of the peak channel temperature in a photoluminescence measurement near the gate overhang on the drain side would be ~ 80 K; a comparable Raman measurement would underestimate the peak by ~ 20 K, based on the smaller spot size of the laser and objective used for Raman measurements.

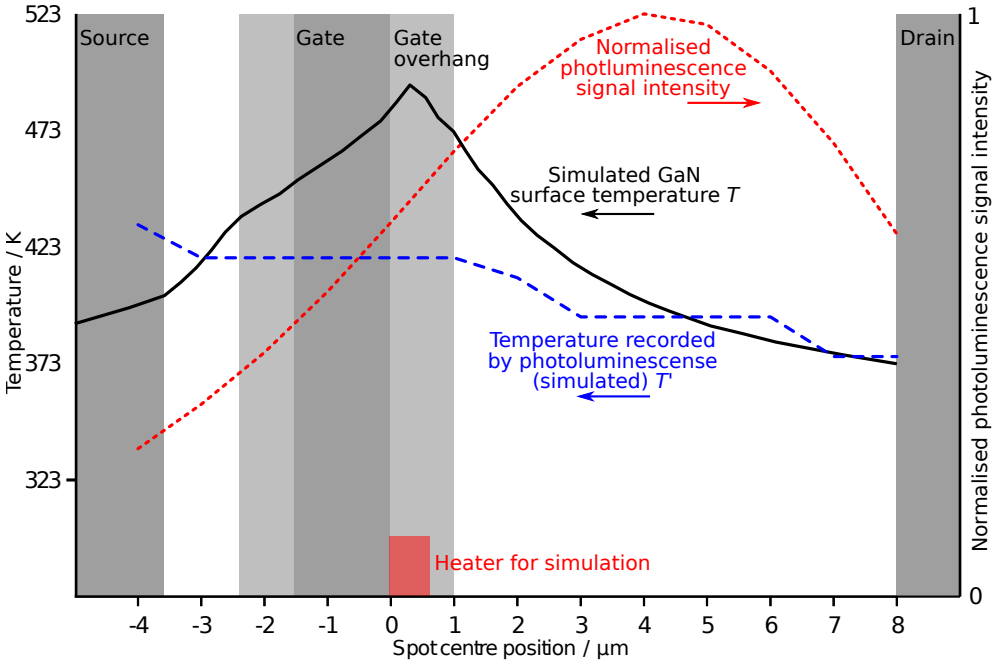


Figure 3.6: Simulated local temperature T , simulated photoluminescence-observed temperature T' (blue dashed line) and simulated photoluminescence signal intensity as the photoluminescence laser is translated over the sample. The device temperature was simulated for a power dissipation of 8.35 W mm^{-1} and a backplate temperature of 25 °C. The photoluminescence laser spot full width at half maximum was 7 µm. The device contacts and gate overhang (shaded) obscure the laser, hence the variation in signal intensity.

3.4 Low temperature measurements and temperature calibration of phonon mode shift

To enable accurate temperature measurements it was necessary to calibrate the phonon mode shift with temperature for each of the device layers. This calibration involved heating or cooling the sample to a series of temperatures over a range such as 83 K to 673 K, in steps of no more than 50 K using a *Linkam THMSG 600* liquid-nitrogen-cooled hot/cold cell (Figure 3.7). At each temperature, three Raman spectra were recorded on an area of the sample free of metal contacts. The resulting spectra had their peak positions extracted using curve fitting function provided by Renishaw within WiRE; for the device discussed in Chapter 4, the peak pick routine developed for use with this layer structure was used. The Cui formula (Section 2.2.5) for the phonon frequency as a function of temperature was then fitted to the phonon peaks for Raman thermography; for photoluminescence thermography the Varshni equation (Equation 2.47) was fitted to the band edge peak. The results of the curve fit for the modes of interest in one sample are shown in Figure 2.13, where the Cui formula is discussed in more detail; the equivalent results for photoluminescence and the fits to them are shown in Figure 2.22.

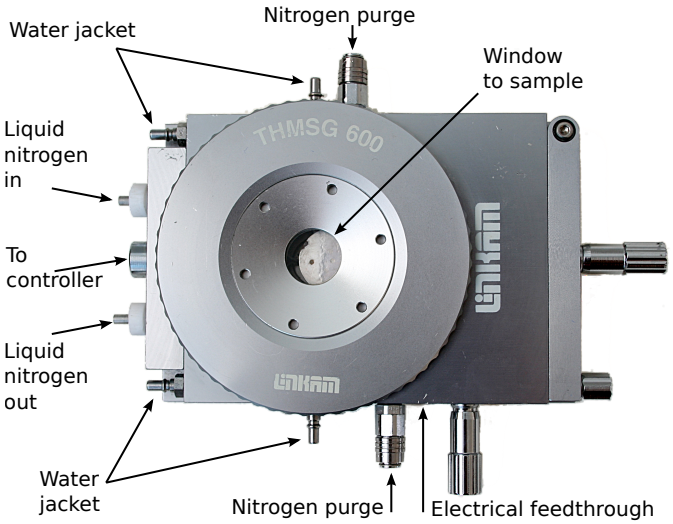


Figure 3.7: The Linkam THMSG 600 hot/cold cell used for temperature calibration of the phonon mode and band-edge photoluminescence peak shifts and for low-temperature electroluminescence. The water jacket prevents the outer surfaces becoming too hot when used at high temperatures, while the nitrogen purge is used to avoid condensation or frost forming on the sample at low temperatures.

The low temperature electroluminescence measurements discussed in Section 6.2.3 were also carried out in the Linkam hot/cold cell. As these measurements required the device to be operated at low temperatures, electrical connections were made by wire-bonding the sample to terminals that could then be connected to the electrical feedthrough of the Linkam cell.

3.5 Thermal simulations

The use of simulations to study the thermal properties of semiconductor devices is well-established, with one-dimensional numerical models being used in the 1970s.[111] As the computing power in a single modern desktop computer is orders of magnitude greater than in the most powerful systems available then,[112] simulations of devices in three dimensions have become a convenient tool for device designers and researchers. Commercially available software means that modelling a device has become a routine task requiring patience but little or no programming. Simulations are used in device design, but also in reliability studies, for example to extract the channel temperature underneath a metal contact (where it cannot be directly observed) from temperatures measured by Raman thermography.[76] Because simulations model the three-dimensional distribution of temperature within a device, they may be used to determine the temperature as a function of depth within an epilayer[56] which requires certain assumptions to be made

about the properties of the sample. Testing these assumptions and comparing simulated with measured depth profiles is discussed in Chapters 4 and 5.

The simulations used here were of the *finite difference* type, in which a large number of coupled differential equations are solved in an iterative numerical process. The software used was *Thermal Analysis System (TAS)* from *ANSYS, inc.* (formerly Harvard Thermal). The model underlying a simulation is a *mesh* made up of *nodes* (infinitesimal points at which temperatures are defined) connected by *resistors*, which are the paths along which heat can flow. It is necessary to optimise the node placement to some extent, as the model convergence time is larger for more nodes, but too few nodes leads to a poor representation of the temperature within the device. Nodes may be placed manually if exact positioning is required, for example on the boundary between two materials, or where they must be placed close together, such as near a source of heat. Alternatively they may be placed automatically, which is appropriate far from the heat source, for example deep in the substrate, where thermal gradients are much smaller. For the models used in this work, the mesh was typically generated automatically, with manual input reserved for regions with a much higher node density than surrounding regions, where the automated routine was prone to inconsistencies. The need for manual intervention in mesh generation was particularly important at the boundaries between regions of high and low node densities, where it was important to ensure that the correct connections were made.

Material properties are assigned to regions of nodes: the main property of interest here is the thermal conductivity (and its temperature dependence). As an example, the room temperature thermal conductivity of typical GaN epilayers is 130 to 160 $\mathrm{W\,m^{-1}\,K^{-1}}$, decreasing with temperature as $T^{-1.4}$. Room temperature thermal conductivities of the substrates relevant to this thesis were given in Table 2.1; they also vary with temperature. *Heater nodes* are placed to simulate the heat dissipation in a device. In a HEMT the peak heat dissipation takes place in a region extending less than 1 µm from the edge of the gate towards the drain, uniformly across the width of the gate.[34] When the device to be simulated also has its temperature measured by Raman thermography, the heater placement may be checked against the experimental data and if necessary adjusted. At the edges of the model, the temperature of the *boundary nodes* may be set, for example to simulate a controlled backplate temperature (Dirichlet boundary condition). The backplate temperature could be related to experimental data by measuring the

substrate temperature far from the device under test under operating conditions, for example ~500 µm, known as the "edge-of-chip" temperature. This was an important step in relating simulations to measurements with high dissipated powers, as the thermal contact between the back of the substrate and the wafer chuck was not perfect and the substrate temperature could rise noticeably. As an example for some of the measurements in Chapter 4 the chuck was at 25 °C while the back of the substrate reached 41 °C. At other edges, symmetry constraints or distance from the heater may allow the assumption to be made that no heat flows across the boundary (Neumann boundary condition). The differential equations coupling the temperatures of the nodes and the heat flow along the resistors are solved iteratively using matrix methods. The iterations continue until the temperature of the nodes changes by an insignificant quantity between consecutive iterations, known as *convergence*.

For a node i the temperature at an iteration j is given by

$$T_{i,j} = (1-\delta)T_{i,j-1} + \delta\frac{\sum_n \frac{T_{n,j-1}}{R_{i,n}} + Q_i}{\sum_n \frac{1}{R_{i,n}}} \tag{3.3}$$

where $R_{i,n}$ is the thermal resistance between nodes i and n, and Q_i is the heat load at node i (which will be zero except at a heater node). The damping parameter δ allows the early stages of the solving process to be sped up by setting $1 < \delta < 2$. As convergence is approached, the value of δ is reduced to less than one, to avoid a situation in which the temperature overshoots the converged value and oscillates about it. Clearly at boundary nodes where Dirichlet boundary conditions were applied $T_{i,j} = T_{i,0}$, and the total heat leaving the model through all such nodes should equal the total heat load $\sum Q_i$ once the model has reached equilibrium – this "heat balance" condition must also be met for the model to be regarded as having converged.

Part of the mesh for a simulation of the device structure studied in Chapter 4 is shown in Figure 3.8. The resulting temperature distribution is shown in Figure 3.9. This model consisted of ~83 000 nodes, and took several hours and hundreds of thousands of iterations to solve initially. A strength of this type of simulation is that small changes to the parameters of a model (for example to the thermal conductivity of a material) typically converge much more quickly than the original model, allowing the values of such parameters to be fitted to measured results. The use of this technique to find the thermal

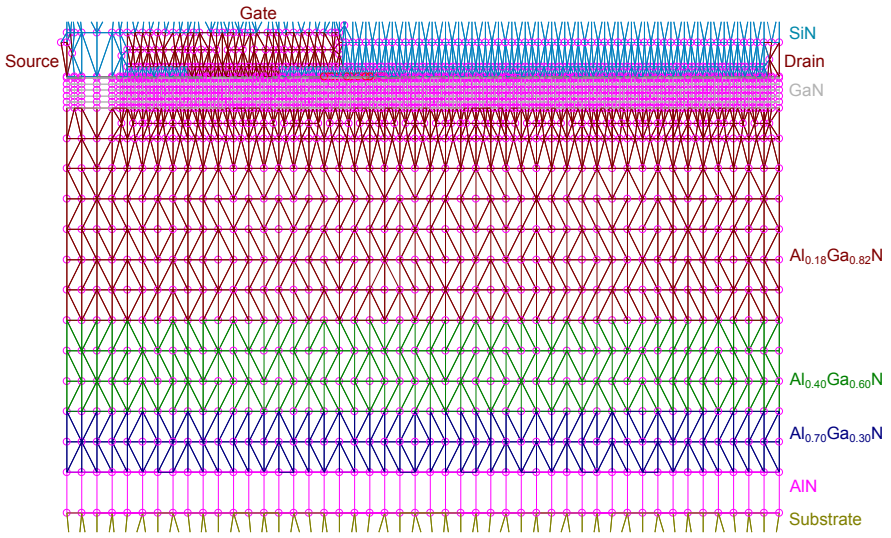


Figure 3.8: Part of the mesh for a simulation of an AlGaN/GaN/AlGaN DHFET (as studied in Chapter 4). Pink circles indicate the location of nodes, and lines show paths along which heat can flow and equations must be solved (resistors). The red nodes are heater nodes. The colours of the lines indicate different materials.

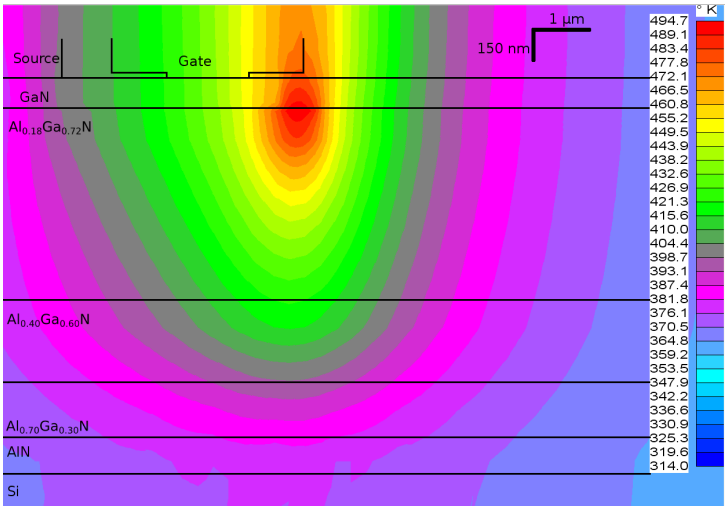


Figure 3.9: The temperature output of the simulation shown in Figure 3.8, for approximately the same region of the model.

conductivity of a thin GaN layer given a temperature profile in the device measured using Raman thermography is described in Chapter 4.

It might be thought that a two-dimensional model consisting of a slice through the centre of the gate, orthogonal to the gate would suffice for simulations. While this may be true for an infinitely long gate, a gate of even 100 µm could not be approximated as infinite therefore a three-dimensional model was always used. If only the heat leaving a slice of the channel perpendicular to the gate was considered, as in a two-dimensional model, the channel temperature would reach unrealistically high values.

3.6 Implications for thermography and reliability studies

The equipment discussed so far allowed the device temperature to be studied in two lateral dimensions, with data from the substrate and epilayers. For transparent substrates (SiC and sapphire) the temperature could be mapped through the substrate, though the resolution worsened significantly as the penetration depth increases. Studying the vertical flow of heat in a device has therefore required simulations, which were constrained by the Raman measurements, meaning that there was an opportunity to develop a more direct means of accessing more temperature data and measuring thermal boundary resistances.

Electroluminescence imaging has always allowed off-state current paths to be located, but has contributed little to our understanding of their nature or formation. A method for observing the development of emerging hotspots in both current and electroluminescence emission should therefore provide a route to understanding their formation. Photoluminescence has a long history of studying semiconductor crystals, and therefore needed only to be applied appropriately to degraded devices to be informative.

Chapter 4

Thin film AlGaN/GaN high electron mobility transistors and vertical temperature profiling

The aim of the work discussed in this chapter was to develop a spectroscopic method for studying the temperature as a function of depth and the heat flow through a novel multi-layer HEMT. The 150 nm GaN channel layer in this structure was embedded in a stack of $Al_xGa_{1-x}N$ layers on a Si substrate. First the device is described in detail, before the experiment and simulation are discussed. The results which follow show that the thinness of the GaN layer leads to a reduced GaN thermal conductivity, which is shown to match theoretical predictions, and has an impact on heat transfer and ultimately device design considerations.

This final result and much of the Raman spectroscopy in this chapter has been published in Reference 113: C. Hodges et al. 'AlGaN/GaN field effect transistors for power electronics – Effect of finite GaN layer thickness on thermal characteristics'. *Applied Physics Letters* 103.20 (2013). The theoretical approach developed by J. Anaya Calvo and discussed here was included in the same paper. The major results were also included (with results from the next chapter) on a conference poster: C. Hodges, J. Anaya Calvo and M. Kuball. *Improving depth resolution of Raman thermography on GaN HEMTs: Temperature in GaN channel and temperature gradients in the GaN buffer layer*. Bristol: UK Nitrides Consortium, Jan. 2014.

4.1 Properties of the devices studied

In power HEMTs, especially those for use at high voltage,[115] vertical breakdown through the stack of epilayers and punch-through of electrons from source to drain through the

GaN buffer layer[116] are significant issues. Vertical breakdown can be reduced by using a thicker buffer combined with optimised buffer doping, such as carbon, or by using an AlGaN back-barrier. An AlGaN back-barrier ensures the the confinement of electrons within the GaN channel,[117] reducing both vertical breakdown and punch-though.[118] The effect of the back barrier can be seen in Figure 4.1(b). This construction may be used to produce a steeper rise in the conduction band level below the channel with respect to depth when compared to an all-GaN buffer, thus improving the confinement of electrons to the buffer. In the devices studied here, the $Al_{0.18}Ga_{0.82}N$ back barrier was 1.8 µm thick, acting as the buffer and forming part of a stack of AlGaN layers which also provided a strain-relief role. The resulting devices were therefore examples of a double heterostructure field effect transistor (DHFET) structure.

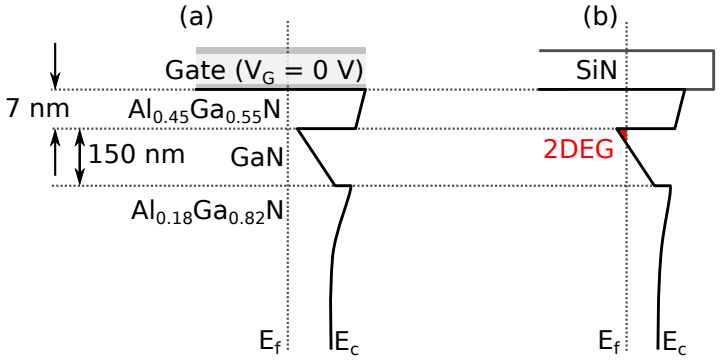


Figure 4.1: Conduction band in the AlGaN/GaN/AlGaN devices: (a) under the gate, with the in-situ SiN removed; (b) in the access region, with in-situ SiN. Figure based on Reference 119.

The layer structure of the devices studied in this chapter is detailed in Figure 4.2. The devices were supplied by S. Stoffels (IMEC). The AlGaN/GaN/AlGaN DHFETs on Si substrates studied here consisted of a 150 nm GaN channel layer on top of multiple $Al_xGa_{1-x}N$ layers, with the aluminium fraction x decreasing in steps from the substrate to the device channel, and with a 7 nm $Al_{0.45}Ga_{0.55}N$ barrier layer on top of the GaN channel. The source-drain spacing was 11.5 µm with a 1.5 µm gate and a gate–drain spacing of 8 µm. The ohmic contacts were TiAlMoAu, and the Schottky gate was Ni/Au. The gate contact was a T-gate with an overhang of 1 µm towards the drain. The devices were operated up to V_{DS}= 20 V, and consisted of two fingers, each of 100 µm gate width, with a gate separation of 101.5 µm. Some properties of the materials used in these devices are given in Table 4.1. The devices studied were enhancement mode HEMTs; e-mode operation was achieved by selective removal of in-situ-grown SiN before gate deposition.[119]

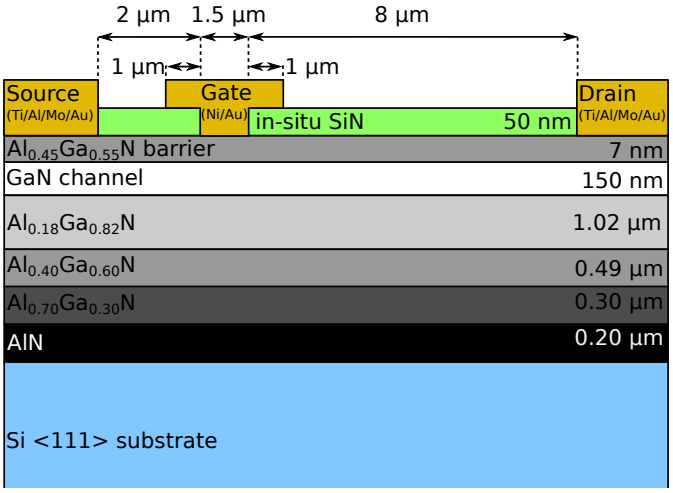


Figure 4.2: The layer structure of the AlGaN/GaN/AlGaN DHFET studied.

4.1.1 Normally-off operation

Enhancement mode or e-mode transistors are preferred in many power applications.[119] One proposed method for achieving this is the gate injection transistor[128] which uses a p-GaN or p-AlGaN gate to raise the conduction band above the Fermi level under the gate. An alternative method for raising the conduction band level above the Fermi level is to use a very thin – less than around 7 nm – AlGaN barrier combined with an AlGaN back-barrier, which has a similar effect on the band structure, raising the conduction

Table 4.1: Room temperature properties of the layers in the device studied in this chapter. The thermal conductivity of GaN varies with temperature as $TC_{\mathrm{GaN}} \propto T^{-1.4}$;[75] the temperature dependence of the thermal conductivity for $Al_xGa_{1-x}N$ and AlN was taken to be the same. For Si the dependence was taken to be $TC_{\mathrm{Si}} \propto T^{-1.3}$.[120]

Material	Lattice constant [a] Å	Lattice structure	Thermal expansion K^{-1}	Thermal conductivity [b] $W\,m^{-1}\,K^{-1}$
GaN	3.19 [c]	Wurtzite	5.6×10^{-6} [c]	160[d]
$Al_{0.18}Ga_{0.82}N$	3.17	Wurtzite		24
$Al_{0.40}Ga_{0.60}N$	3.16	Wurtzite		14
$Al_{0.70}Ga_{0.30}N$	3.13	Wurtzite		17
AlN	3.11 [c]	Wurtzite	4.2×10^{-6} [c]	285[e]
Si (111)	3.84 [f]	Diamond Cubic	2.6×10^{-6} [g]	149[h]

[a] The lattice constants for GaN, $Al_xGa_{1-x}N$ and AlN are given for the *a*-axis. For $Al_xGa_{1-x}N$ they are calculated using Vegard's law.[28,29]
[b] Values for AlGaN from reference 121 based on a model in reference 122.
[c] From reference 123.
[d] Bulk value, from reference 75.
[e] From reference 124.
[f] From reference 125.
[g] From reference 126.
[h] From reference 127.

band level to just above the Fermi level under the gate.[129] The role of the AlGaN back-barrier is to cause negative polarisation charge to build up at the interface between the GaN and the back-barrier, depleting the channel to the extent that a positive gate voltage is required for current to flow. In the devices studied in this chapter, an in-situ-grown SiN cap was used to prevent surface depletion of the 2DEG.[130] This allowed the conduction band minimum to stay below the Fermi level under the SiN. By selectively removing this SiN under the gate, the AlGaN surface potential was modified so that the conduction band level was raised above the Fermi level under the gate, depleting the 2DEG as grown. This effect may be seen in Figure 4.1. The enhancement mode design led to a pinch-off voltage of +475 mV.

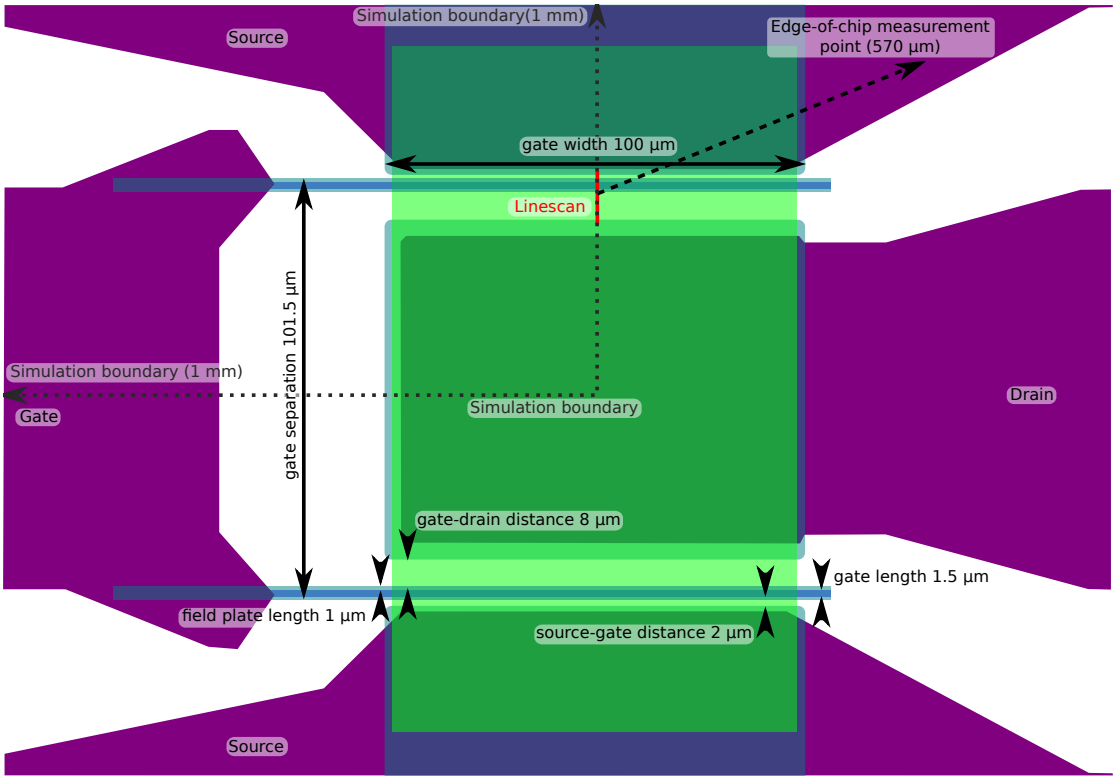


Figure 4.3: Layout of the thin film AlGaN/GaN/AlGaN DHFET, showing the linescan measurement position and the boundaries of the simulation, which were taken to be lines of symmetry in the heat flow within the device assuming that both fingers dissipate the same power. The blue areas are metal contacts (including the gate overhang), the purple areas are metal interconnects, and the green shows the location of an additional SiN layer. Based on a figure provided by S. Stoffels, IMEC.[131]

4.1.2 Thermal implications

The thermal conductivity of $Al_xGa_{1-x}N$ is generally lower than that of either GaN or AlN.[132,133] As the majority of the heat is generated at the drain edge of the gate[34,100] in the GaN device channel and must be extracted through the buffer and the substrate

to the heatsink on the back of the substrate, the reduced (compared to GaN) thermal conductivity of the AlGaN buffer will have an adverse effect on the peak channel temperature. The relatively high thermal conductivity of the GaN channel layer on top of the AlGaN buffer will thus cause it to act as a heat spreader. This is generally true for all structures in which there is an AlGaN layer in between the GaN channel and the heat extraction path to the substrate, for example N-polar devices.[21] The thermal resistance of the AlGaN layers is in addition to the extra thermal boundary resistances which will be caused by the additional interfaces in a multilayer device compared to a conventional HEMT which has only a single thin nucleation layer between the GaN buffer and the substrate. Measuring the effect of these phenomena on the peak channel temperature is the purpose of the work described in this chapter.

4.2 Raman Spectroscopy of AlGaN layers

4.2.1 Phonon modes and Al-fraction – implications for thermography

The phonon modes in $Al_xGa_{1-x}N$ which are commonly used for Raman thermography, namely the E_2 (high) and A_1, both shift in frequency with the aluminium fraction. This has been studied by Liu *et al.*,[134] Demangeot *et al.*,[135] and most completely by Davydov *et al.*[136] The A_1 mode shifts in a simple manner to higher frequencies at higher aluminium fractions, showing broadening of the Raman peak at intermediate compositions due to alloy disorder. The E_2 mode displays more complex behaviour – in addition to broadening and shift with Al-fraction it splits into AlN- and GaN-like branches, between which the energy is shared in approximate proportion to the Al-fraction.[136] These effects are illustrated in Figure 4.4.

The measured frequencies of these lines for the compositions relevant here are given in Table 4.2, compared with those previously reported. Differences between the predicted and measured band positions may be attributed to strain in the AlGaN layers – the role of these layers is to relieve strain, due to both thermal and lattice mismatch, so they will necessarily be strained themselves. As mentioned above, the E_2 Raman mode splits into AlN- and GaN-like branches in AlGaN. Both modes show an increase in frequency with Al-fraction, but the rate of change with Al-fraction for the AlN-like mode is highest for small Al-fractions, and for the GaN-like mode it is highest for large Al-fractions. The

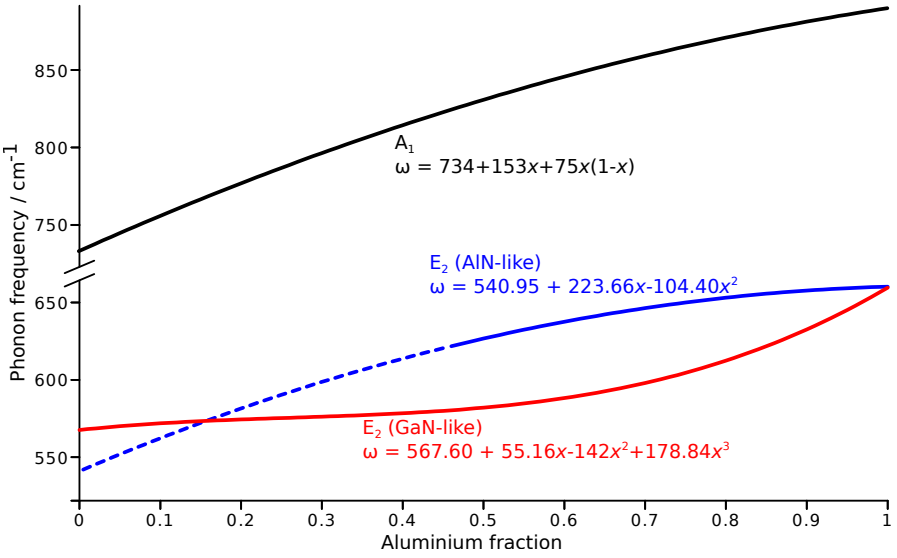


Figure 4.4: The phonon frequency predicted as a function of aluminium fraction for the phonon modes used in Raman thermography. The dashed line shows the continuation of the curve predicted in Reference 135, over a range for which no experimental data was observed. Data for the A_1 mode was taken from Reference 136.

effect of this is that the AlN-like modes can be observed in the Raman spectrum, while the GaN-like modes instead combine to form a shoulder to the much stronger GaN E_2 Raman peak. The AlN-like E_2 mode from the $Al_{0.18}Ga_{0.82}N$ also contributes to this shoulder.

As each layer, i.e. each aluminium fraction, offers both an A_1 and an (AlN-like) E_2 peak, it may be expected that all the layers can be probed. Figure 4.5 shows a Raman spectrum of the device studied here, obtained under more favourable measuring conditions than were feasible for mapping the device temperature as a function of position. From this figure it is clear that, quite apart from the combination into a shoulder of the GaN-like E_2 lines, by no means all of the predicted lines listed in Table 4.2 can be resolved with sufficient precision for accurate curve fitting. The signal-to-noise ratio in

Table 4.2: Room temperature Raman band positions calculated from literature[135,136] and measured here for the phonons in the GaN, $Al_xGa_{1-x}N$ and AlN layers present in the AlGaN/GaN/AlGaN DHFET sample.

	x (%)	0	18	40	70	100
A_1 (LO)	predicted	746	777	813	858	891
	measured		781	823		
E_2 (GaN-like)	predicted	568	574	578	598	660
	measured	572				
E_2 (AlN-like)	predicted	541	578	614	646	660
	measured			618	647	

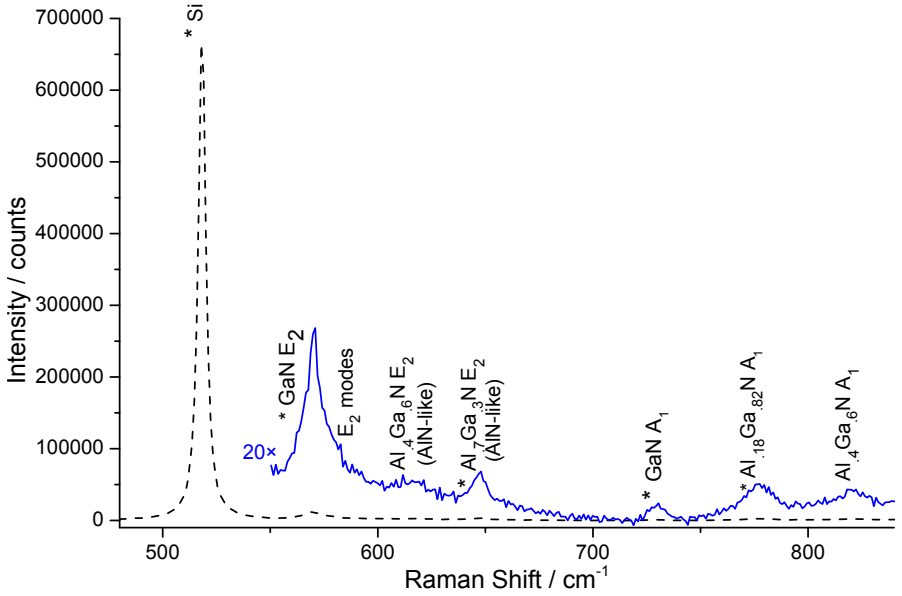


Figure 4.5: Raman spectrum from the AlGaN/GaN/AlGaN DHFET (an average of five spectra recorded outside the device at room temperature). The shoulder labelled "E_2 modes" consists of the GaN-like E_2 lines from all the AlGaN layers, as well as the AlN-like E_2 line from the $Al_{0.18}Ga_{0.82}N$ layer. Modes used in thermography are indicated by asterisks.

this sample was adversely affected by a rather high background (particularly in the region of the rather weak A_1 peaks). In addition not only the splitting of the E_2 lines and the broadening mentioned above due to disorder[137,138] in AlGaN but also broadening due to differences in strain within a layer weaken the signal from the AlGaN. In fact although the $Al_{0.40}Ga_{0.60}N$ layer had two resolvable peaks in test spectra such as that shown in Figure 4.5, neither line could be located with any usable degree of precision in an experimentally-feasible timescale, especially given the possibility of sample-stage drift (both thermal and due to vibration). Effects due to the depth of focus of the laser spot also contribute to the relative weakness of the peaks from lower layers compared to the GaN channel, which was nearest the laser focus at the sample surface, while the $Al_{0.18}Ga_{0.82}N$ layer, at 1.8 µm thick, occupies most of the rest region of maximum laser intensity and system response (Section 2.3). The GaN layer may also be assumed to have good crystal quality, enhancing the Raman signal compared to the AlGaN layers, as the stepped AlGaN strain relief layers were designed to terminate threading dislocations at the interfaces between them.[139] The net effect of this was to reduce the signal strength from the deeper layers in comparison to the GaN and $Al_{0.18}Ga_{0.82}N$, a reduction increased by the fact that these layers are thin compared to the $Al_{0.18}Ga_{0.82}N$ buffer. Similarly no Raman peak could be measured from the AlN nucleation layer, despite the long acquisition times used here (100 s per position) and the thickness of the layer compared to typical HEMTs, as this layer is far from the laser focus and AlN nucleation

layers exhibit poor crystal quality due to the lattice mismatch with the substrate,[140] which would tend to weaken the Raman signal from the nucleation layer.

4.2.2 Detecting Raman peak positions

Due to the complexity of the Raman spectrum from these devices, care had to be taken to correctly determine the Raman peak positions. For samples giving simpler spectra, it is typically possible to extract the peak positions using Renishaw's *WiRE* software, provided with their Raman microscopes. The curve fit routine in *WiRE* allows fitting of single or multiple peaks (Gaussian, Lorentzian or mixed) automatically, extracting the peak position, peak height, and other parameters. Constraints can be placed on the parameters of each peak, and a baseline can automatically be fitted and subtracted. This routine often fails, however, in spectra such as the one shown in Figure 4.5, where a weak peak of interest is close in frequency to a much stronger peak, or when superimposed on a non-flat background. Curve fitting routines in general attempt to minimise the residual between the input data and the fitted curve, expressed as the sum of the squares of the differences between each data point and the corresponding position in the function being fitted.[141] Attempting to fit weak peaks adjacent to strong peaks (for example in this case the weak GaN E_2 Raman peak next to the strong Si peak) can therefore cause residuals in the strong peaks to dominate the overall residual curve with the result that the weak peak is not reliably fitted. Similarly the shape of the background may resemble a broad peak and dominate the residuals, causing the fitting function to attempt to fit the background instead of the peak, as was observed for the AlGaN A_1 peaks from this sample. While it is possible in principle[142] to fit the largest peak first, and fix the parameters of that peak before fitting the rest of the spectrum, the *WiRE* software does not make this possible in processing map data.

An alternative method was therefore required to extract the peak positions from the spectra. The general approach to first-derivative peak-picking follows that described by O'Haver,[142] from which the process described below was adapted.

The concept of peak-picking is rather simple. At the location of a peak in the spectrum, the first derivative of the spectrum will pass through zero. In a Raman spectrum, where only positive peaks are of interest, only those zero-crossings in the first derivative that cross from positive to negative are of interest. To make this work in practice, in a noisy

spectrum with a spectral step size comparable to the precision with which a peak is to be located, a few more steps are required, as shown in Figure 4.6. *Smoothing* of the original spectrum is required to reduce the effects of noise. As is common in spectroscopy the Savitzky-Golay[143] method was used. In the absence of *interpolation*, the zero-crossings in the spectrum can only be found to a precision equal to the spacing of points in the spectrum. In practice this meant that precision for a given peak position was very limited – less than ten unique positions for each peak over the temperature range observed in a device – therefore the spectrum was interpolated onto a finer grid using cubic spline interpolation.

It may be desirable to carry out a *background spectrum subtraction,* especially if peaks are situated on sloping regions of background in the Raman spectrum. The background spectrum must have good signal-to-noise ratio so as not to introduce additional noise and make the process of peak-picking more difficult. In practice this meant averaging many background spectra collected under identical conditions matching the measurement conditions. Whether this is beneficial or not depends on the sources of the background. If the background originates from the sample, it is not possible to collect a pure background spectrum, but if the background is due to optical components of the spectrometer, a spectrum collected on a null sample (such as a gold mirror) may be used as a background. Background subtraction was not found to be helpful on the thin layer sample studied here, as most of the background appeared to come from the sample itself. However the same method was used in analysing some of the other experiments discussed elsewhere in this thesis, when subtraction of an instrument-related background was beneficial.

After interpolation a further smoothing step was carried out as the Savitzky-Golay method was used once again to take the derivative of the spectrum.[141] The derivative was then scanned to find the points where it crossed through zero from positive to negative, which were then listed as the positions of the Raman peaks in the spectrum. These steps can be seen in Figure 4.7, showing data with good signal-to-noise ratio and therefore few spurious peaks. In realistic spectra there will almost always be zero-crossings in the spectrum caused by noise, even after smoothing. As many of the Raman peaks from this sample were broad and weak, and therefore the magnitude of their first derivative was small, it was often the case that some of spurious zero-crossings were near peaks of interest.

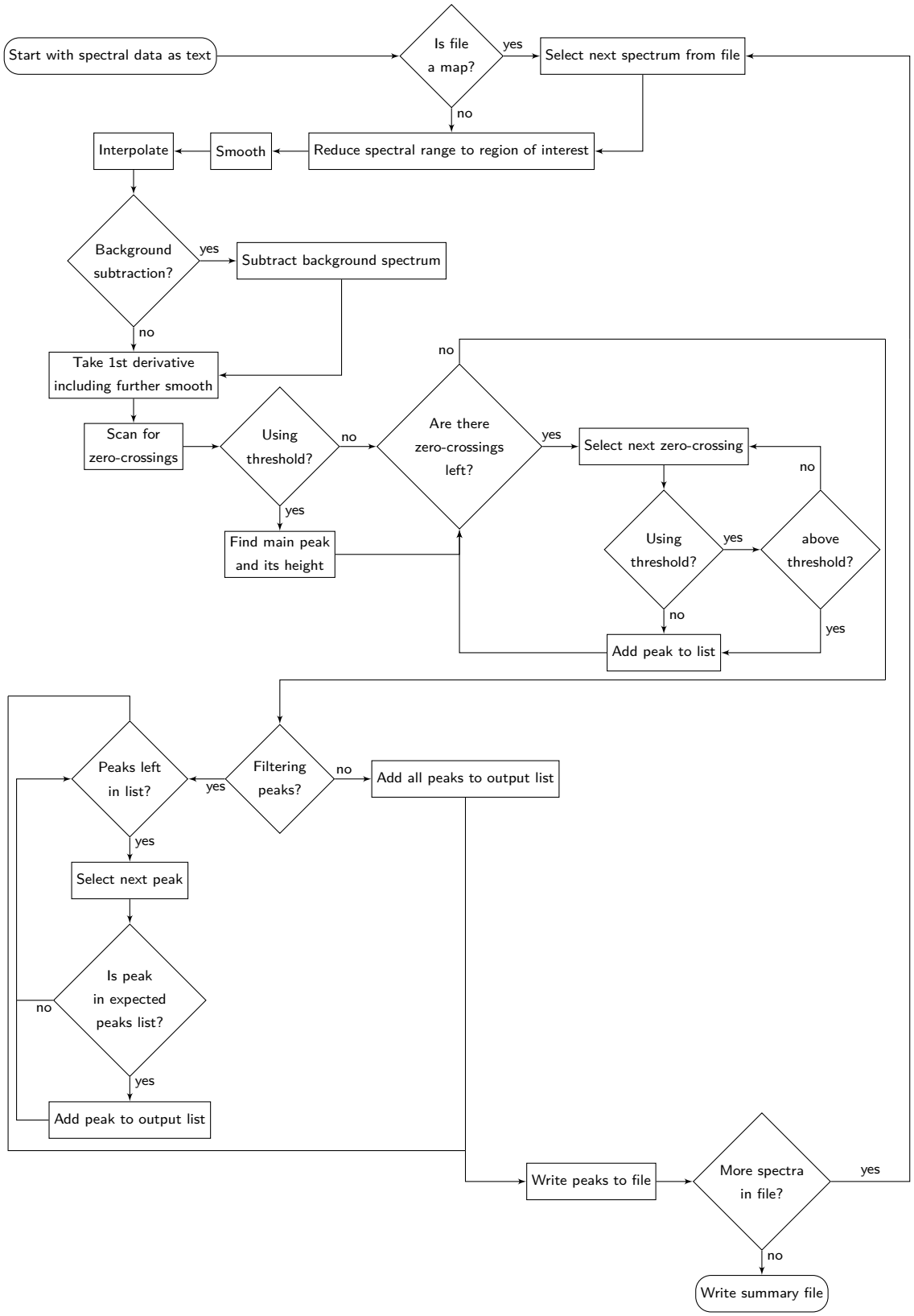


Figure 4.6: Simplified peak-picking process for data in a single file. The user is expected to supply the positions of the peaks for filtering; the threshold level and main peak; and any background spectrum, which is smoothed and interpolated in the same way as the spectra under analysis.

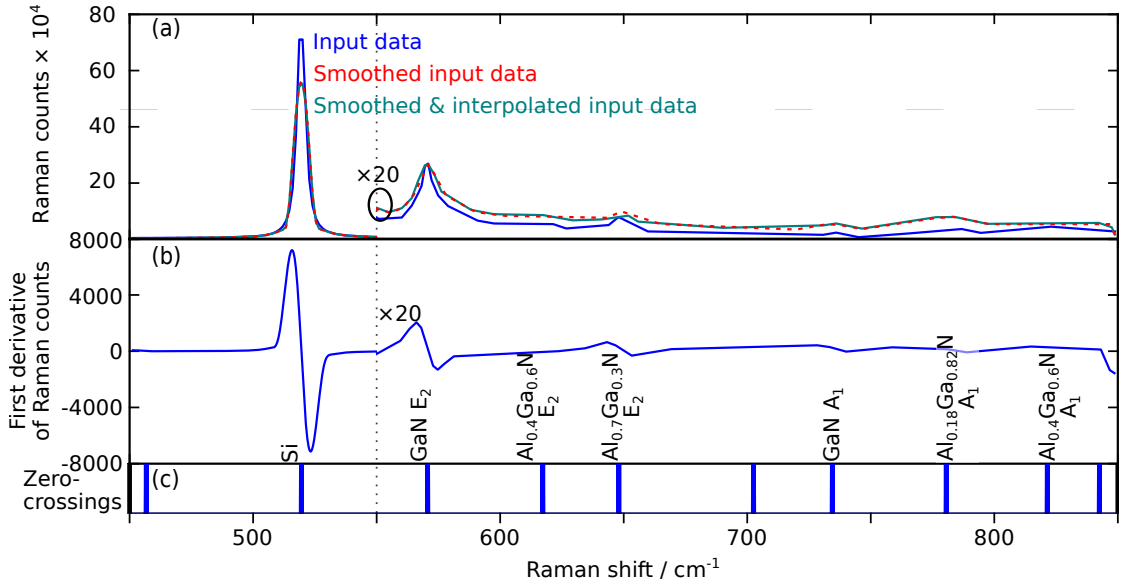


Figure 4.7: Illustration of the peak-picking process for detecting Raman peak positions: (a) the input Raman spectrum and the effect of smoothing and interpolation on that data; (b) the first derivative of the spectrum, showing the zero-crossings corresponding to the peaks in the Raman spectrum; (c) the positions of the zero-crossings detected in the derivative, indicating those assigned to Raman bands. For this spectrum (shown in more detail in Figure 4.5) the few spurious peaks detected could easily be discarded by the "filtering" process described in the main text as being far from any expected peaks.

In the presence of noise (i.e. in any real spectrum) curve fitting is often preferred over peak-picking as many data points contribute to specifying the parameters of a peak, averaging out some of the noise, in addition to allowing sub-pixel location of the peak. This influence of points spread over the whole spectrum is also the cause of failure to fit weak peaks near stronger peaks or to fit on a complex baseline. In the peak-picking method described here far fewer data points contribute to the finding of a peak (the exact number and contribution being determined by the smoothing parameters), reducing the accuracy with which it can be located, but cubic spline interpolation, by taking into account the derivatives of the spectrum near a peak, allows it to still be located more precisely than simply using a local maximum in the uninterpolated data. The uncertainty in the peak position may be larger than with curve fitting, but the use of multiple acquisitions allowed this uncertainty to be reduced and its effect on the temperature results to be quantified.

To avoid falsely identifying peak positions it was necessary to remove spurious peaks from the output of the peak-picking routine. Two methods were used: First a simple *threshold* in the first derivative discarded peaks smaller than a configurable fraction of the largest (typically substrate-related) peak. Then the peaks found were *filtered* by comparing to a list of expected peaks – only those within a user-specified window centred

at an expected peak position were accepted. In practice it was necessary to change these expected peak positions during the thermal calibration. Considering the peak position shift with temperature over the large temperature range necessary for calibration, a large window would have had to be used otherwise, and may spurious peaks would have passed through the filter. This was not necessary for the analysis of the linescan data used to obtain a device temperature profile, as the temperature range in the device, and therefore the range of positions for a given Raman peak, was much more limited than in the calibration.

If any spurious peaks remained after these two processes, a manual approach was required: the magnitude of the first derivative could be compared between multiple matches for an expected peak; the spectrum could be inspected; or curve fitting of a single peak could be used to select between ambiguous matches to a single expected peak. In practice for some peaks all three methods had to be used. In some cases all the candidate matches had to be discarded as none could be confidently deemed to be correct. The peak-picking code is given in Appendix A.1.

4.2.3 Two-dimensional temperature mapping and extraction of the peak temperature

A Raman linescan across the entire drain–gate access region was used to obtain a temperature profile within the devices. In this case the devices were scanned under the laser spot in 0.25 µm steps, recording temperature Raman spectra at each step. A 532 nm laser was used with a spot size of 0.5 µm and a numerical aperture (NA) of 0.5. The temperature of a given layer of the device was determined from the shift of the layer's Raman peak from the corresponding spectrum in a reference linescan recorded with the device pinched off. The results were then combined with a simulation (Section 3.5) to extract the peak temperature, taking into account the lateral spatial averaging inherent in the finite laser spot size.

The layer structure present in this device allowed a two-dimensional map to be produced from a one-dimensional linescan, by using the spectroscopic addressability of the varying AlGaN composition. Multiple linescans were acquired, so that for each peak at each location an average peak position could be found, improving the signal-to-noise ratio and allowing calculation of the uncertainty in the results. The A_1 phonon mode was used

where possible to determine the device temperature (for the GaN and $Al_{0.70}Ga_{0.30}N$ it was necessary to use the E_2 line for improved signal-to-noise ratio in the temperature measurement). When collecting linescans of 0.25 µm step size over the source-drain distance of 11.5 µm the stability of the system limited the maximum acquisition time to 100 s per point. With this acquisition time, especially close to the metal contacts, the signal-to-noise ratio was not sufficient to measure the position of all the bands identifiable in Figure 4.5, which was measured several microns away from the nearest metal contacts.

The Raman signal intensity as a function of position in the linescan between source and drain, passing over the gate is shown in Figure 4.8. Illustrated is the the Si substrate peak intensity, a similar intensity profile may be assumed for the weaker epilayer peaks. Effectively this meant that it was not always possible to locate the peak, and therefore probe the temperature, for all of the epilayers in the device as the measurement point approached the contacts. The reduced signal-to-noise ratio for the $Al_xGa_{1-x}N$ Raman peaks also led to higher noise in the measured temperature data for these layers compared to the GaN and the Si substrate. In particular, no band was reliably measurable for the $Al_{0.40}Ga_{0.60}N$ layer, so no experimental temperature for this layer was available.

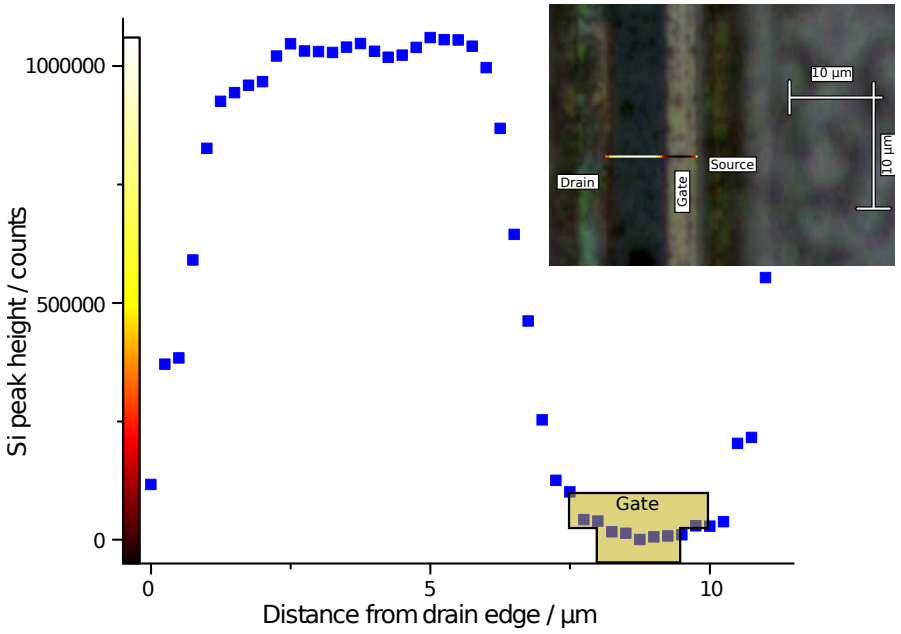


Figure 4.8: Si substrate peak intensity between drain and source on the AlGaN/GaN/AlGaN DHFET. The inset shows the same data overlaid on a white-light image of the device, with the peak height shown by a colour scale.

4.3 Photoluminescence thermography

It is possible to probe the temperature of a GaN surface by means of photoluminescence, as discussed in Section 2.4. This was carried out on the devices studied here to measure the GaN channel temperature and provide a consistency check on the Raman measurements. Photoluminescence thermography is a surface-sensitive technique, which on typical GaN buffers would be used to measure the temperature of the top of the GaN, in contrast to the depth average inherent in Raman thermography. In the DHFET structure which is the subject of this chapter, the GaN channel layer is thin enough that the Raman and photoluminescence thermography probe a similar depth. There is, however, a significant difference: photoluminescence has a worse lateral spatial resolution than Raman, as discussed in Section 3.3, therefore the results must be interpreted with care on a non-isothermal sample, such as a HEMT. The photoluminescence temperature dependence was also calibrated for this specific wafer, in a similar way to the calibration of the phonon mode shifts for Raman thermography, as described in Section 3.4.

4.4 Thermal model

In order to find the peak device temperature, which is typically located under the gate overhang or near the edge of the gate,[34,100] where optical access is challenging, it was necessary to fit a thermal model to the experimental data. This thermal model was created using *Thermal Analysis System* finite difference software as described in Section 3.5. The backplate temperature was adjusted to match the "edge-of-chip" temperature, measured 570 µm from the linescan, which was in the centre of the device finger as shown in Figure 4.3.

As Raman thermography measures average temperature over each device layer thickness,[56] the temperature in the simulation was averaged over a volume corresponding to the laser focus within each layer when comparing to experimental data, using a weighted average in the $Al_{0.18}Ga_{0.82}N$ and a simple average in the thinner layers. The thin device layers showed essentially no temperature gradient with depth in the simulation, but the $Al_{0.18}Ga_{0.82}N$ layer, at 1.8 µm thick and with a low thermal conductivity, did show a significant temperature gradient, making this an important step. The lateral dimension of the laser spot within the sample was also taken into account.

4.5 Results and discussion

4.5.1 Raman results of device temperature

Figure 4.9 shows the temperature distribution in the AlGaN/GaN DHFET in the GaN channel and the different AlGaN layers in the gate-drain access region of the device, with the device operated at 6.5 W mm^{-1}. The temperature peaks near the drain side of the gate contact in the GaN channel. This is where the majority of the electrical power is dissipated as Joule heating in the device. The temperature decreases towards the Si substrate, as is apparent with each AlGaN layer closer to the Si substrate, with the heat sink of the device being at the back of the Si substrate. This figure demonstrates that the presence of different compositions of $Al_xGa_{1-x}N$ therefore provides the possibility to access three-dimensional thermal information of benefit when subsequently comparing experimental to thermal simulation data. This is of use when extracting the thermal conductivity of the GaN epilayer, as will be seen in Section 4.5.3.

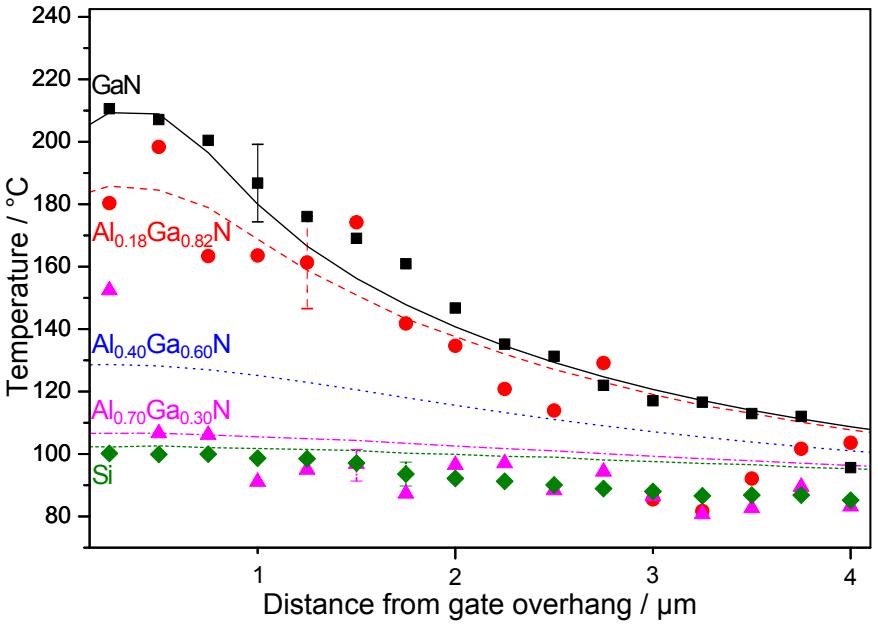


Figure 4.9: Experimental (points) and simulated (lines) temperature distribution of an AlGaN/GaN/AlGaN DHFET across the gate-drain access region, with the device operated at $V_{DS} = 20$ V, $V_{GS} = 1$ V ($I_D = 325$ mA mm^{-1}) at 6.5 W mm^{-1} with a backplate temperature of 25 °C. Representative error bars are shown.

The relationship between dissipated power in the device and the device temperature profile was explored, with particular attention to the peak channel temperature. Individual Raman spectra were acquired at the drain side of the gate at increasing electrical powers.

The devices were operated at drain-source voltages, V_{DS} of 2 V to 20 V and a gate-source voltage V_{GS} of +1.6 V, giving power densities of 0.55 W mm^{-1} to 8.4 W mm^{-1}. The device temperatures obtained, for the GaN channel and silicon substrate only, are shown in Figure 4.10, along with simulation data for the same power densities. The channel and substrate temperatures can be seen to increase faster than linearly with power ($T_{GaN} \propto P^{1.38}$, $T_{Si} \propto P^{1.59}$). This has been seen in previous work,[75,144,145] and as the thermal conductivities of the materials in the device decrease with increasing temperature, is to be expected. However the simulated temperature as a function of power shows a less steep increase ($T_{GaN} \propto P^{1.18}$, $T_{Si} \propto P^{1.32}$) using the model calibrated to the temperature distribution at 6.5 W mm^{-1} as shown in Figure 4.9 despite a good match to the experimental data at both low and high powers. One possible reason for this difference is a possible shift in the peak electric field with drain bias, and therefore in the location of the temperature peak.[100] The location of the peak heat dissipation will be discussed in Section 4.5.3.

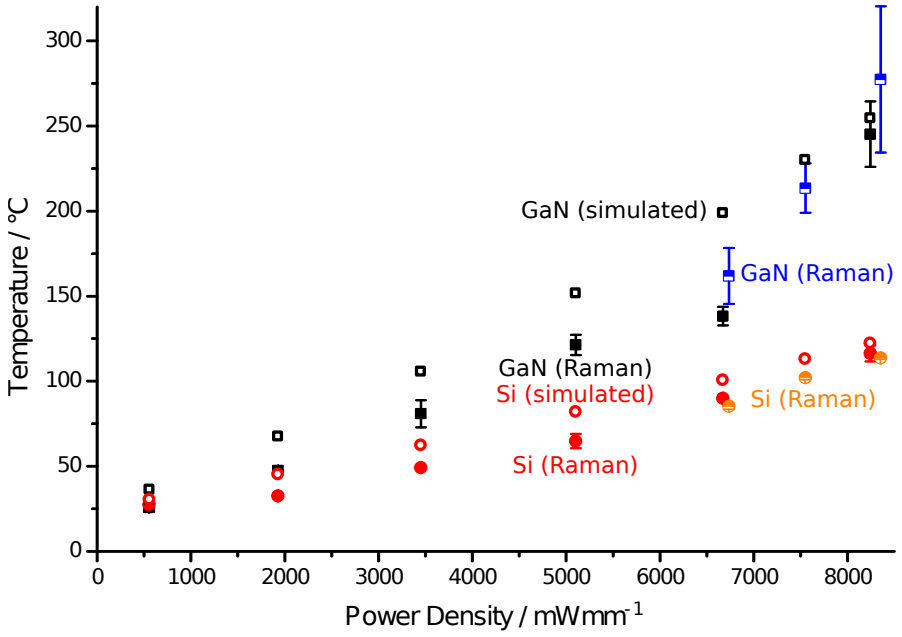


Figure 4.10: Peak temperature dependence of the GaN channel (squares) and Si substrate (circles) on the electrical power density in the device. Filled and half filled symbols show temperatures measured by Raman spectroscopy at the drain edge of the gate. The blue and orange half filled symbols show Raman data collected at a later date, and averaged over fewer acquisitions, than the black and red filled symbols. Open symbols show simulated temperatures.

4.5.2 Photoluminescence results

The GaN channel temperature in the device was measured using photoluminescence spectroscopy. As the photoluminescence measurements used different measurement conditions (a single finger was operated, at similar power densities, resulting in lower total powers) simulations matching these conditions were carried out.

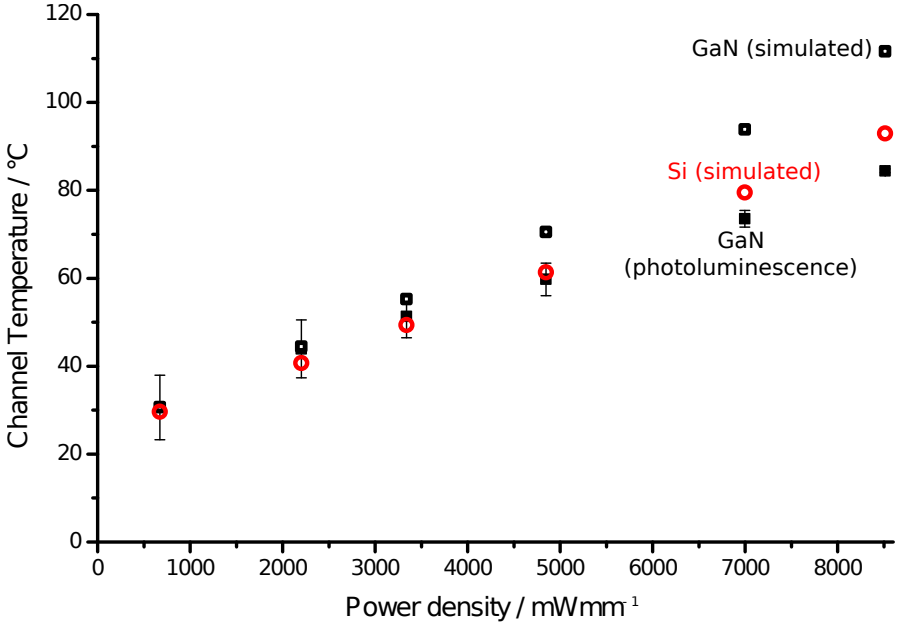


Figure 4.11: GaN channel temperature measured by photoluminescence spectroscopy near the middle of the drain-gate access region, as a function of power (solid squares) and simulated (open squares). The open red circles show the simulated Si substrate temperature. A single finger was operated for the PL experiment.

Recalling that the photoluminescence laser spot size is around 5 µm due to the much lower objective numerical aperture than for Raman thermography and the chromatic aberration between the white-light focus and the laser focus, the GaN temperature was averaged over a region greater than the whole region shown in Figure 4.9, thus averaging over a significant temperature gradient. Figure 4.11 illustrates this, showing a much lower channel temperature than Figure 4.10. This is not, however, the only cause for the reduced peak channel temperature found by PL. Some of the difference can be explained by the fact that in the two-finger operation used in Raman thermography measurements, no heat could flow from the finger being measured towards the other finger, whereas in the PL measurements, as the other finger was unpowered, an extra path for heat loss was enabled. The simulated data shown in Figure 4.11 takes both of the effect of spatial averaging and the additional heat flow into account, but still gives a higher temperature than measured by photoluminescence. The analysis of the simulation data

does not however take into account the asymmetry in the PL peak which is likely to affect the peak spectral position when PL spectra from regions of varying temperatures are superimposed. Quantifying this effect is not simple given that it is interdependent with both the spatial averaging and changes in peak height with temperature. It may however be estimated for a simple test case based on spectra recorded at known temperatures for the purpose of calibration. An unweighted average of spectra recorded at 298 K and 373 K would ideally have a peak corresponding to 335.5 K. If such a spectrum is generated using calibration spectra, the band edge peak is found to be at a position corresponding to 328 K. This demonstrates that the error introduced by averaging over the peak shape, for the temperature range of interest, will tend to bias temperature results downwards.

4.5.3 Thermal modelling, location of heat in the channel and nanoscale thermal effects

As the simulation was purely thermal, self-heating was not incorporated directly in the model. Instead, based on previous work,[100] the thermal simulation used a heater 0.6 µm long near the gate overhang, with a constant power density over the full width of the gate. The power density was set to match the total dissipated power in a Raman or photoluminescence experiment to which the simluation was to be compared. It is to be expected that most of the Joule heating takes place in the peak field region, and that the peak field is near the edge of the gate, as has previously been reported based on electrical simulations,[34] and electroluminescence.[100] This assumption was tested to ensure that it provided the best agreement of the simulated to the experimental data. The effect of the position of the heat dissipation in the simulation is shown in Figure 4.12, in which it can be seen that if the heat source was located under the gate foot, or split between the gate foot and the gate overhang (which may both be the location of significant electric fields) the temperature profile in the GaN channel measured by Raman spectroscopy could not be matched by the simulation. The results of the simulation for all layers can be seen in Figure 4.9 along with the experimental data to which it was matched. Initially, literature values for thermal conductivity were used for all layers. This led to a considerable underestimate of the peak temperature in the GaN channel and a flatter temperature distribution in the GaN away from the gate contact than observed, whether the GaN thermal conductivity was that of good-quality modern epitaxy (160 $\mathrm{W\,m^{-1}\,K^{-1}}$) or of

older epitaxy (130 $\mathrm{W\,m^{-1}\,K^{-1}}$).[146] Therefore the dependence of the temperature profile on the thermal conductivities of the layer structure was investigated. It was found that reducing the substrate, AlN, $Al_{0.70}Ga_{0.30}N$ and $Al_{0.40}Ga_{0.60}N$ thermal conductivities in the simulations had little effect on the peak channel temperature or the channel temperature profile. As these are below the significant barrier to heat flow presented by the thick, low thermal conductivity $Al_{0.18}Ga_{0.82}N$ buffer, it is not surprising that their effect will be small. Decreasing the thermal conductivity of the $Al_{0.18}Ga_{0.82}N$ buffer only had a significant effect on the peak channel temperature and not the lateral temperature distribution. This is shown in Figure 4.12, where (compared to both the experimental results and the simulation with a reduced GaN thermal conductivity) the GaN temperature is higher, with a broader distribution, for a GaN thermal conductivity suitable for a thick epilayer and a reduced $Al_{0.18}Ga_{0.82}N$ thermal conductivity. Considering the already low thermal conductivity of AlGaN, this is to be expected, as most of the temperature drop occurs across the $Al_{0.18}Ga_{0.82}N$, and the GaN channel layer acts as a heat spreader in this type of device. Varying the thermal conductivity of the GaN channel, while using literature values for the $Al_xGa_{1-x}N$ and Si thermal conductivity was therefore the only way to obtain the best fit to the temperature profile observed experimentally.

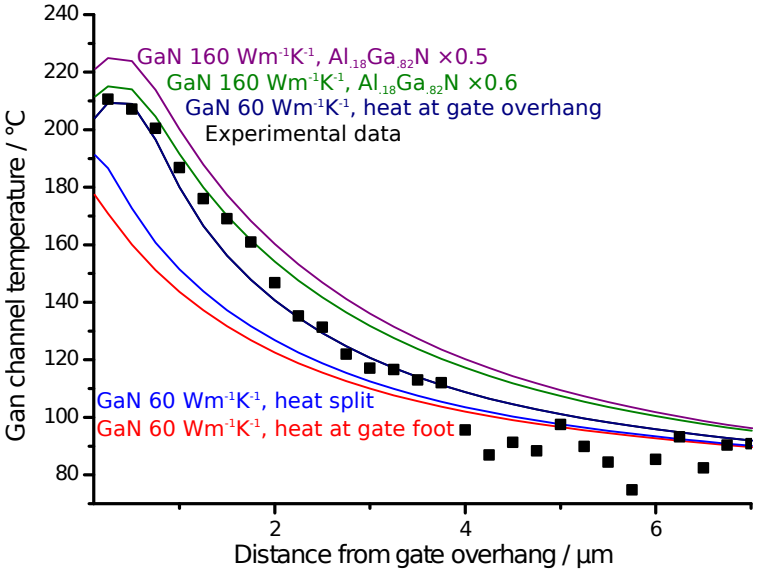


Figure 4.12: The effect of varying the heat position and the thermal conductivity of the $Al_{0.18}Ga_{0.82}N$ buffer on the GaN channel temperature profile (lines), along with the experimentally measured GaN channel temperature (squares).

The thermal conductivity of the GaN channel was found to be (60 ± 20) $\mathrm{W\,m^{-1}\,K^{-1}}$ by changing the thermal conductivity of the GaN; extracting the temperature profiles within the device;[a] and calculating the root-mean-square (RMS) difference between the

[a]Including experimentally-equivalent spatial averaging as described in Section 5.2.3.

experimental and simulated temperature distributions in the access region, for all layers except the $Al_{0.40}Ga_{0.60}N$. This is significantly lower than typical good quality modern GaN which may be expected to have a thermal conductivity of 160 $W\,m^{-1}\,K^{-1}$. Figure 4.13 compares the simulated temperature profile for a range of GaN thermal conductivities with the experimental GaN temperature profile.

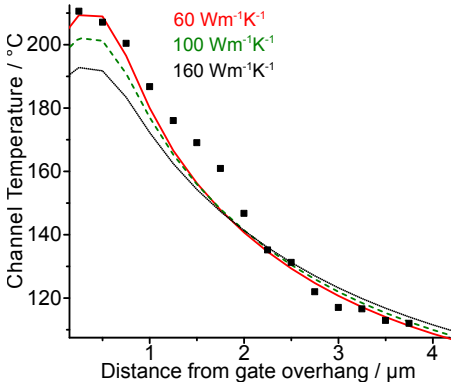


Figure 4.13: A comparison of the simulated (lines) and experimental (squares) temperature profiles for the GaN channel in an AlGaN/GaN/AlGaN DHFET, for a range of simulated thermal conductivities.

This reduced GaN thermal conductivity required further consideration. A first-principles model of heat flow in thin films was applied to this GaN layer. The approach used was originally proposed for thin silicon films,[54] and was applied in more detail to this sample by J. Anaya Calvo,[113] resulting in the range of thermal conductivities shown in Figure 4.14 for the range of reported sound velocities in GaN.[147–149] In nanostructures, the mean free path of the phonons is comparable to the thickness L of the layer,[150–152] as thermal transport is dominated by the scattering events at the boundaries. This means that heat transport due to phonons can be considered to be a form of Knudsen flow,[151,152] as the (bulk) mean free path (MFP) Λ_B becomes similar to the layer thickness l in this regime, and therefore the Knudsen number $K_n = {}^{\Lambda_B}/_{l}$ approaches 1. The thin film thermal conductivity, k_{Film}, can be related to that of the bulk, k_B by

$$k_{Film} = k_B \frac{\Lambda_{eff}}{\Lambda_B} \tag{4.1}$$

where the effective MFP, taking into account boundary scattering, is related to the Knudsen number by $\Lambda_{eff} = {}^1/_2(\Lambda_1 + \Lambda_2)$ and[54]

$$\begin{aligned}
\Lambda_1 &= \Lambda_B K_n \int_0^{\frac{1}{K_n}} \left(1 + (\alpha - 1)e^{-\alpha} - \alpha^2 \int_a^{\infty} \beta^{-1} e^{-\beta}\, d\beta\right) d\alpha \\
\Lambda_2 &= \Lambda_B \left(1 - \left(\frac{1}{K_n} - 1\right) e^{\frac{1}{K_n}} - \left(\frac{1}{K_n}\right)^2 \int_{\frac{1}{K_n}}^{\infty} \beta^{-1} e^{-\beta}\, d\beta\right)
\end{aligned} \tag{4.2}$$

where α and β are integration variables with no physical meaning. The bulk phonon MFP can be found classically from $k_B = {}^1/_3 C_v \Lambda_B v$, where C_v is the heat capacity per unit volume and v is the sound velocity. A range of average sound velocities from 3338 $\mathrm{m\,s^{-1}}$ to 5000 $\mathrm{m\,s^{-1}}$ have been reported for GaN.[148] Combining these with the accepted bulk GaN thermal conductivity of 160 $\mathrm{W\,m^{-1}\,K^{-1}}$, specific heat of 490 $\mathrm{J\,kg^{-1}\,K^{-1}}$ and density of 6150 $\mathrm{kg\,m^{-3}}$ allows the range of phonon MFPs to be found to be 32 nm to 48 nm, and from Equation 4.1 the thermal conductivity to be determined to be 62 $\mathrm{W\,m^{-1}\,K^{-1}}$ to 78 $\mathrm{W\,m^{-1}\,K^{-1}}$, in excellent agreement with the result obtained above by fitting the simulated thermal conductivity to the device temperature profile measured by Raman thermography.

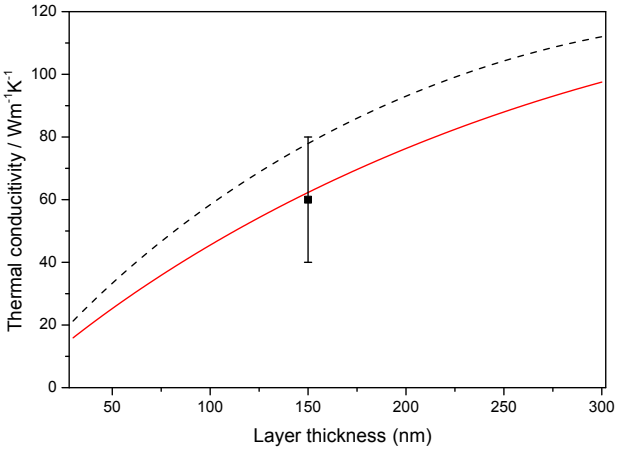


Figure 4.14: Modelled GaN thermal conductivity dependence on thickness for sound velocities at 300 K of 3338 $\mathrm{m\,s^{-1}}$ (solid red curve) and 5000 $\mathrm{m\,s^{-1}}$ (dashed black curve). The experimental value determined here is shown by a square. Figure and model by J. Anaya Calvo in Reference 113.

From Figure 4.13 it can be seen that the reduced thermal conductivity of the GaN channel causes an additional temperature rise of around 10% compared to that of bulk GaN. This in addition to the already elevated temperature due to the low thermal conductivity of the AlGaN layer stack, and is likely to adversely affect the reliability of devices based on thin-film structures similar to the one discussed here. It will be necessary to take this factor into account in the design of such power HEMTs, especially as the layer thickness is reduced still further,[153] as can be seen from Figure 4.14. While similar effects may take place in thin strain-relief layers and even nucleation layers, they are likely to be indistinguishable from the low thermal conductivity due to crystal quality in these layers and the inherently low thermal conductivity of AlGaN compared to GaN. In many cases the effect will be included in the effective thermal boundary resistance at the buffer-nucleation layer-barrier interface for layers other than channel layers.

4.6 Conclusions

Raman spectra from e-mode AlGaN/GaN/AlGaN DHFETs on a Si substrate were analysed to extract data corresponding to the characteristic phonon frequencies of the different AlGaN compositions present as buffer and strain relief layers. A method was developed for automated peak-picking of the rather cluttered spectra from such samples. The temperatures of these layers were measured, in addition to the GaN channel layer temperature, enabling a three-dimensional temperature map to be produced. Simulations were used to extract the thermal conductivity of this 150 nm GaN channel from the Raman thermography data. The thermal conductivity of the thin channel layer was found to be $\sim 60\ \mathrm{W\,m^{-1}\,K^{-1}}$, which is significantly less than that of a typical GaN epilayer ($\sim 160\ \mathrm{W\,m^{-1}\,K^{-1}}$) and is a good match to theoretical predictions due to the confinement of the phonon MFP in a thin layer. Due to the inherent low thermal conductivity of the AlGaN device layers the GaN acts as a heat spreader in the devices. This thermal benefit of the GaN channel layer is however reduced in this thin layer, and would disappear if its thickness were reduced to the tens of nanometres range. The resulting increased peak channel temperature which this causes has implications which need to be considered for optimising AlGaN/GaN/AlGaN DHFET designs.

With the current drive for AlGaN/GaN HEMTs on Si substrates, particularly for high voltage operation, layer structures similar to the one studied here are likely to become more common. The approach discussed here is generally applicable to such devices, with the caveats that there must be sufficient optical access and the compositions of the layers must be sufficiently different for the peaks to be resolved. The latter caveat is unlikely to be an issue, but the former suggests that experimental devices will be required, as there is a real need for field plates on high-voltage devices, which will obscure optical access to the channel. In the presence of field plates which obstruct a significant fraction of the channel, the extra constraints placed on a thermal model by temperature data from multiple layers within the device mean that multi-layer Raman thermography may still make a useful contribution to understanding such devices.

Chapter 5

Improving the confocality of Raman systems to probe temperature gradients in AlGaN/GaN HEMTs

The goal of the experiments discussed in this chapter was to measure the temperatures of the top (i.e. buffer-barrier interface) and bottom (i.e. buffer-substrate interface) of a GaN buffer in a HEMT. In addition measuring the thermal boundary resistance between the buffer and the substrate was envisaged. The approach taken was to modify a standard Raman microscope system to improve its optical depth resolution. This is desirable to place extra constraints on thermal models of devices, as well as to provide independent support for simulation results.

The main results of this chapter, as exemplified by Figure 5.16, along with some of the supporting work from Chapter 2, were published in Reference 154: C. Hodges, J. Pomeroy and M. Kuball. 'Probing temperature gradients within the GaN buffer layer of AlGaN/GaN high electron mobility transistors with Raman thermography'. *Journal of Applied Physics* 115.6 (2014). In addition the main result was included on a conference poster with the results from the previous chapter: C. Hodges, J. Anaya Calvo and M. Kuball. *Improving depth resolution of Raman thermography on GaN HEMTs: Temperature in GaN channel and temperature gradients in the GaN buffer layer*. Bristol: UK Nitrides Consortium, Jan. 2014.

As substrate thermal conductivity increases, more and more of the temperature drop in a device – from the surface where heat is generated to the heatsink at the back of the substrate – is across the GaN buffer (assuming the thermal interface resistance to the substrate remains constant). Therefore the thermal gradient across the GaN buffer in the devices is steeper. Self-heating is highly localised both in its lateral position at the drain edge of the gate and in depth, at the buffer-barrier interface. Existing Raman

thermography techniques provide a high lateral resolution, enabling mapping of the lateral temperature distribution (as was seen in Chapter 4), but the depth resolution remains limited to measuring a single buffer temperature except in devices such as those discussed in Chapter 4 which have spectroscopically addressable layers due to different AlGaN compositions. In devices with a significant temperature drop across a homogeneous buffer this will lead to an underestimate of the measured peak temperature, an important parameter in thermal degradation; typically thermal simulations are used to extract the channel temperature from the measurement data. When measuring from the back of an optically transparent wafer, for example to measure underneath metal contacts or field plates, depth resolution suffers even further, therefore any improvement in the optical depth resolution would be beneficial, even if it only improves the resolution of a thermal depth map in the substrate without giving access to additional epilayer information.

Azimuthal and radial polarisations are explored in this work to remove the effects of birefringence when focussing through the optically anisotropic materials typically used in AlGaN/GaN HEMTs, and to detect the E_1 phonon mode using a geometry which does not normally permit this. Two methods for generating azimuthal and radial polarisation are explored: (i) exploiting the two foci formed by rays refracted by the different refractive indices present in a birefringent material; (ii) the use of a spatially-varying waveplate based on nanoscale structures.

5.1 Modifications to the Renishaw micro-Raman system

As was discussed in Section 2.3, the depth resolution available for a measurement depends on the diffraction limit and refraction (i.e. spherical aberration) at the sample surface. Equation 2.33 (repeated here for reference)

$$w_{\mathrm{ax}} = \sqrt{\left(\frac{0.88\lambda}{n-\sqrt{n^2-NA^2}}\right)^2 + \left(\frac{\sqrt{2}nPH}{NA}\right)^2 + \frac{\ln 2}{2}\Delta^2\left(\left[\frac{NA^2\left(n^2-1\right)}{1-NA^2}+n^2\right]^{1/2}-n\right)^2}$$

gave the full width at half maximum of a Gaussian approximation to the combined axial spot size w_{ax} of a confocal microscope system at a wavelength λ due to refraction and diffraction in terms of the refractive index, n, of the medium in which the focus is formed, the numerical aperture (NA) of the focussing lens, the diameter of the confocal

pinhole imaged into sample space (PH), and Δ, the position of the nominal focus below the sample surface. These variables were shown in Figure 2.16.

This approximation is valid for small values of Δ, but for applications such as measuring from the back of the wafer with an objective lens which lacks variable coverslip correction it would be necessary to calculate the convolution of the effects of refraction and diffraction to assess the depth resolution. Figure 5.1(a) plots the depth resolution w_{ax} for a range of microscope objectives focussing from air into GaN. The 63× objective shown was used for most of the experimental work in this chapter. A 100× 0.9 numerical aperture objective is included in Figure 5.1(a) as a notional upper limit for dry (i.e. non-immersion) objective lenses. To do better than this would require oil immersion in addition to the other factors explored in this chapter. Neither a 100× 0.9 numerical aperture lens nor an oil-immersion lens would be likely to allow the device under test (DUT) to be contacted with probe tips given the short working distances these lenses typically have.

Treating the standard Renishaw micro-Raman system (as shown in Figure 3.2) as a confocal microscope is an approximation, but a reasonable one. A true confocal microscope passes the excitation light through a pinhole which is confocal with the sample, and passes the return light (both Raman-scattered and reflected) back through the same pinhole before analysing it. In the standard configuration of the Renishaw system, the excitation laser may optionally be passed through a pinhole in the beam expander. The standard configuration also uses a *virtual pinhole* formed by the spectrometer's adjustable entrance slit (which also determines the spectral resolution) and, orthogonal to the slit, the CCD image area.[155] This determines the spatial resolution in place of the confocal pinhole assumed for Equation 2.33. With the modifications discussed in this chapter, the system becomes a confocal microscope system as a confocal pinhole is inserted.

To find w_{ax} for the standard Renishaw system from Equation 2.33 it is necessary to define PH for the virtual pinhole. To do this it is first necessary to project the CCD image area onto the slit, then to project this effective CCD area and slit width onto the sample. The slit width and effective CCD area are usually matched to each other for Raman point measurements and maps; for other measuring techniques such as photoluminescence and electroluminescence spectroscopy a larger image area may be used.

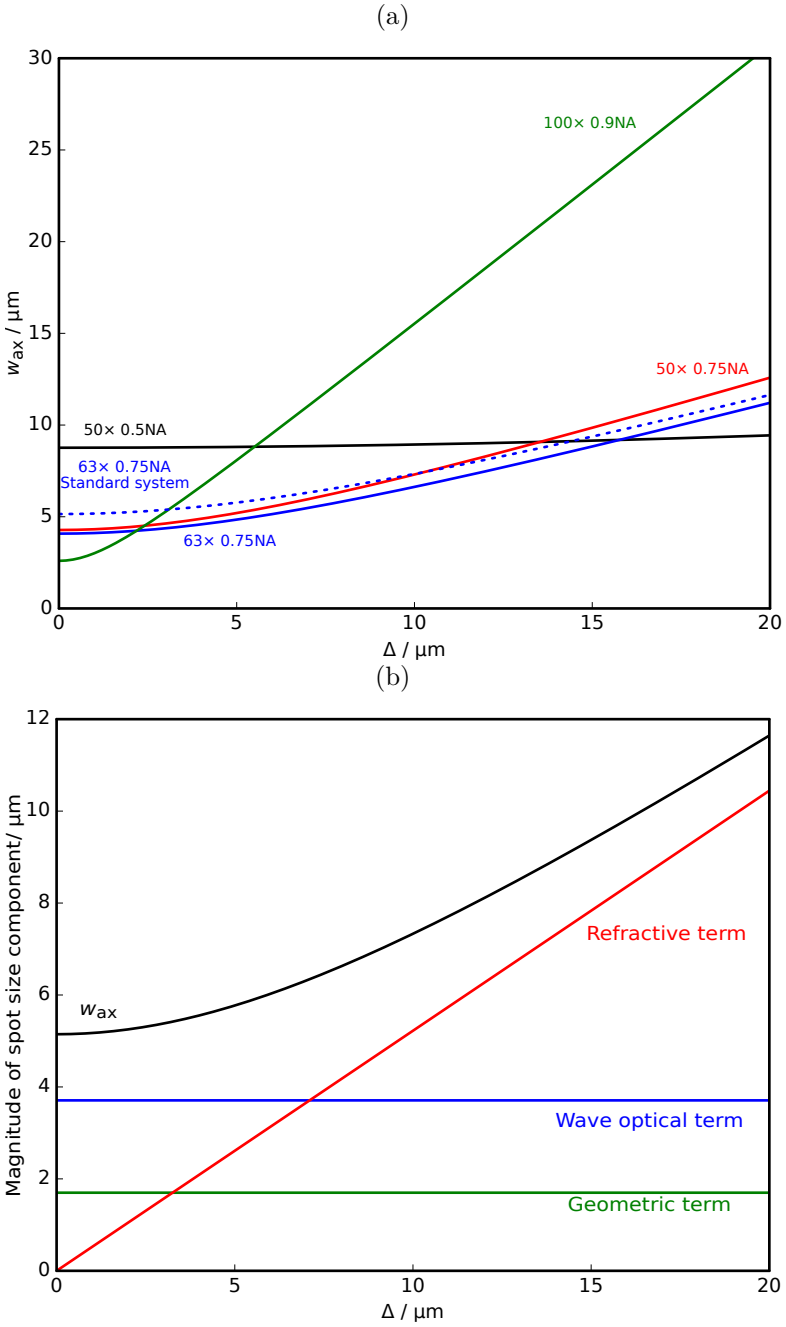


Figure 5.1: (a) The axial resolution as a function of depth in GaN calculated from Equation 2.33 for a range of microscope objectives. For the solid curves the pinhole and lenses are as shown in Figure 5.4(a), the dotted curve compares the effect of the standard system pinhole. The refractive index $n_{\mathrm{GaN}} = 2.39$. (b) The relative importance of the three terms in Equation 2.33 for the standard system with the 63× lens as a function of depth.

Recalling from Section 3.1.1 that a CCD column consists of light at a particular wavelength, while the CCD rows differ in their spatial coordinates, it is possible to optimise the system confocality by minimising the size of the virtual pinhole. Maximum confocality – the smallest axial diffraction limit – is obtained by using a microscope objective with a high numerical aperture and a short focal length combined with a one-pixel-high column on the CCD, and setting the spectrometer entrance slit width to match the size of a single pixel imaged at the the slit. The CCD pixels on the Renishaw micro-Raman system used in this work are 22 µm, which is magnified by a factor of $^3/_5$ onto the slit, so if a single pixel is selected this is equivalent to a slit width of 13.2 µm. The slit width is set by hand using a micrometer, and for high-confocality measurements is set between 10 µm and 20 µm. It is therefore a reasonable approximation to consider an effective pinhole size of 15 µm. To a good approximation Equations 2.25 and 2.33 determine the spatial resolution that can be obtained in thermography measurements. Figure 5.1(a) illustrates the effect of changing the objective and inserting a confocal pinhole on the depth resolution of the system over a range of depths.

Even in the absence of a confocal pinhole, the spectrometer spatial response of a typical Raman microscope system (defined by the slit and CCD image area) is designed to be matched to the size of the laser spot. This helps to optimise an experiment's signal-to-noise ratio, by reducing the contribution from stray light and from fluorescence or Raman scattering caused by the tails of the laser spot, which may fall on the device contacts or similarly unwanted regions. Matching the CCD image area to the laser spot size also reduces the number of pixels that have to be read out, decreasing the background (due to dark current, stray light and laser-generated light from near the measurement point) and the associated shot noise and readout noise.

To decrease the axial spot size it is necessary to decrease one or more of the three terms in Equation 2.33, where the greatest effect will come from reducing the largest term – which term this is will depend on the experiment, in particular the objective *NA* and the depth Δ. Below the GaN surface ($n = 2.3$) the first two terms are similar in magnitude, while the third is depth-dependent, and dominates for $\Delta \gtrsim 7$ µm, as shown in Figure 5.1(b).

The first (wave optical) term depends on the laser wavelength and the numerical aperture of the microscope objective, so a shorter wavelength and a higher numerical aperture will improve the resolution. For practical measurements on operating devices sufficient

working distance – i.e. the distance between the end of the microscope objective and its focal plane – for probe tips is required if measurements are being made from the contact side of the device. DC probe tips limit the minimum working distance to around 1 mm, which excludes most microscope objectives with a numerical aperture greater than 0.75 for visible lasers, or 0.5 for ultraviolet lasers; the available lenses meeting these criteria were listed in Table 3.1. For RF probe tips the required working distance is typically larger, which further restricts the available objective numerical aperture. The lower achievable numerical aperture in the ultraviolet than in the visible indicates that reducing the laser wavelength into the ultraviolet would not be beneficial, at least in terms of resolution. Even within the visible range the choice of laser wavelengths is limited by availability, leading to the use of 532 nm and 488 nm lasers for the work in this chapter. As visible objective lenses have poor transmission below about 400 nm, the resolution obtainable with the 488 nm laser in particular approaches the practical limit.

The second term under the root in Equation 2.33 is a geometric term, and depends mainly on the size of the pinhole, as imaged into sample space. As the standard Renishaw spectrometer system has an effective pinhole diameter of around 15 µm imaged by the combination of the microscope objective and the preslit lens (see Figure 3.2), inserting an additional pinhole into the system, by placing a spatial filter in between the microscope and the spectrometer, was regarded as the best way to improve the axial resolution, despite the loss of signal that this would inevitably cause. Clearly a pinhole cannot be made arbitrarily small, both because light must pass through it for the experiment, and because of diffraction. The pinhole size and the focal lengths of the lenses focussing through it were optimised taking into account these constraints, which will be discussed in more detail in Section 5.1.3.

The third term in Equation 2.33 represents spherical aberration due to refraction at a planar sample surface, and may be taken to be zero if the focus is formed in air, as in this case there is no surface at which refraction can occur. This term is larger for larger values of the objective numerical aperture, in contrast to the other two terms. As Δ and n are fundamental properties of the experiment and sample respectively, there is no simple way to reduce this term, except by reducing the numerical aperture, which would increase the other terms.

As an example, it is useful to consider an experimental setup typical for a thermography experiment. Given that the pre-slit lens of the Renishaw spectrometer has a focal length

of 50 mm, for a microscope objective with a focal length of 4 mm[a] with a numerical aperture of 0.5, the axial spot full width at half maximum in air (with its refractive index of 1) is 4.7 µm, and the radial spot full width at half maximum is 0.5 µm. The microscope objective used when attempting to maximise the confocality of the system was a Zeiss LD Plan-Neofluar 63×, with a focal length of 2.62 mm and a numerical aperture of 0.75.[b] This leads to an axial spot full width at half maximum in air of 2.0 µm, and a lateral spot full width at half maximum of 0.33 µm.

Assuming that when the laser is focussed on the surface of the GaN buffer this can be treated as a focus in air (a best-case scenario which will be illustrated in Figure 5.7) for a typical 1.9 µm buffer the on-axis laser intensity ratio between top and bottom of the GaN is 1.6:1 for the 50× Leica objective and 12:1 for the 63× Zeiss objective, calculated from the Gaussian spot size given by Equation 2.33. This means that the spectrometer can only measure a weighted average Raman band position through the depth of the buffer, with a much more surface-sensitive weighting for the 63× lens. As will be seen in Figure 5.16, there is a significant temperature gradient across a GaN buffer in the on-state. The combination of this gradient with the axial spot size means that a weighted average of the buffer temperature is measured.[56]

If the laser is focussed below the surface of the GaN, such as in an attempt to measure the temperature of layers deeper inside the sample, the spot becomes much larger. The lateral diffraction limit increases due to the change in effective numerical aperture, but the axial diffraction limit given by Equation 2.26 immediately becomes 3.6 µm (for the Zeiss 63× objective) due to the change in refractive index. The effect of refraction at the sample surface is not included in this calculation and can only make that resolution worse. Although the effect of refraction is not dominant in Equation 2.33 for Δ <4 µm in GaN or SiC, it increases the axial spot size significantly for $\Delta \gtrsim 2$ µm even for a well-chosen objective (Figure 5.1(b)). Thus, as may be seen from Section 2.3.2, the effect of spherical aberration is a significant obstacle to obtaining a temperature or strain profile as a function of depth within an AlGaN/GaN HEMT. It is therefore desirable to find methods of minimising or circumventing this loss of resolution, by optical or other means.

[a] A 50× Leica long working distance objective, a typical objective for work on operating devices.

[b] The nominal magnification of this lens is 63×, for use with a tube lens of 165 mm – the Leica 50× lens was designed for use with a tube lens of 200 mm, hence the apparent discrepancy in the relationship between magnification and focal length.

Optically, spherical aberration can be addressed by means of a solid immersion lens (SIL) or by using an objective lens with variable coverslip correction. The Zeiss 63× objective lens used here has variable coverslip correction, but this feature is not of much benefit when the penetration of the laser into the sample is variable (such as when attempting to obtain a vertical temperature profile through a device) as the cover slip correction must be set as a function of penetration depth and the objective must then be refocussed. In an automated depth map this is not possible, and the need for refocussing would be likely to introduce errors in the z-position in a manual approach. However when measuring the buffer though the substrate ($\Delta \sim$100 µm to 300 µm) any fractional change in the thickness over a depth measurement is small compared to the absolute thickness and the uncorrected spherical aberration is so severe that it is highly beneficial to use this correction. Similarly if a solid immersion lens (Section 5.1.2) is used at depths other than those for which it was designed, its benefit is reduced. This is because the rays of the laser beam no longer arrive normal to the surface and are refracted, though by less than at a planar surface. Both of these methods are therefore only really suited for use at a fixed or nearly fixed measurement depths. Despite this limitation the solid immersion lens is discussed in more detail in Section 5.1.2, as results obtained with such an optic will be compared to those obtained using the spatial filter, and as a solid immersion lens was used in generating azimuthal polarisation (Section 5.1.6).

5.1.1 Alternatives to Raman thermography

Raman thermography is of course not the only way to measure the temperature of a semiconductor. In this section some other techniques for measuring the temperature of operating semiconductor structures are discussed, with reference to their ability to provide temperature profiles as a function of depth.

Photoluminescence thermography provides a surface-sensitive measurement and was discussed in Sections 2.4 and 3.3. It does however have a worse lateral resolution than Raman thermography, so while it has been used to constrain the surface temperature in device simulations, it must be used with care and its benefit when measuring the temperature of HEMTs is limited.

Thermoreflectance[156] is sensitive to changes in refractive index with temperature. thermoreflectance gives a similar lateral spatial resolution to Raman thermography, as similar wavelengths are commonly used. Because it relies on reflection from an interface where the refractive index changes it is also surface sensitive, however it is challenging to remove the contribution to the signal from the interface between the buffer and the substrate, and more challenging still to measure just at this interface.

Scanning thermal microscopy is a technique based on atomic force microscopy (AFM) using temperature sensitive probe tips. It is used for both thermal conductivity measurements[146] and temperature measurements.[157] As a contact technique, it is highly surface sensitive, but has issues with calibration (for which Raman thermography may be used), and gives no signal from anything below the surface. The presence of conductive tips in the high electric field present on the surface of HEMTs causes problems for this technique.

Infrared imaging is widely used but has poor lateral resolution compared to Raman thermography.[158] As GaN and its substrates are transparent at the wavelengths used infrared techniques give a signal from a wide and ill-defined range of depths unless the sample surface is coated, in which case the technique is surface-sensitive and damages the sample. It may however be useful as a screening technique to identify devices or regions with abnormal thermal properties.

Electrical methods[159] are useful for measuring the average channel temperature and integrate well with electrical testing. They generally provide no spatial information or substrate temperature, though they may be used to calculate the thermal resistance when combined with simulations. Newer techniques provide access to the temperature under the gate[160] but require specially-prepared samples, making them of limited use in the general case.

Liquid crystal thermography[161] has a long history and good lateral resolution, but affects the thermal and electrical environment of the DUT, as the liquid crystal is placed directly on the device surface. To extract a temperature profile requires a complex and time-consuming experiment and analysis. Once again, this is a surface-sensitive technique providing no depth information.

What these diverse methods have in common is that compared to Raman thermography they provide reduced depth information (i.e. they do not easily give access even to the

temperature of the substrate) therefore to improve the measurement of temperature with depth improving the depth resolution of Raman thermography is a promising approach. There is more than one way to improve the depth resolution of Raman thermography, in fact a combination of methods is likely to give the greatest benefit.

5.1.2 The solid immersion lens

In many fields, immersion microscopy is used to improve the resolution by increasing the refractive index of the working medium. When used for microscopy within a sample the use of immersion media can significantly reduce the effect of refraction at the sample surface.[162] Common immersion microscopy uses water (refractive index 1.33) or immersion oil (refractive index 1.52); an objective must be used with the immersion medium for which it was designed. Specialist high-refractive-index fluids exist with refractive indices approaching that of GaN, but they are toxic, corrosive and have poor transmission at blue and green wavelengths which may be chosen for high spatial resolution. There are also no objective lenses designed for these fluids, and when used near a focus they are likely to lead to unwanted Raman bands and fluorescence in a Raman spectrum. Furthermore, in experiments on electrical devices, the electrical properties of any material added to the device must be taken into consideration – clearly a conductive fluid applied to the contact side of a HEMT would not be appropriate if the device were to have a bias applied. An alternative approach has been developed for high-refractive-index materials – the solid immersion lens.[163] Solid immersion lenses have been applied to semiconductors such as GaP[164] and silicon,[165] for which commercial systems have become available.

The general principle of a solid immersion lens – also known as a numerical aperture increasing lens – is simply expressed as placing a lens on the sample surface such that no refraction takes place where the light (i) enters the lens or (ii) where it passes from the lens into the sample. The first condition requires that all rays are normal to the lens surface, implying a hemisphere centred on the objective focus, as illustrated in Figure 5.2. The second condition requires the the refractive indices of the solid immersion lens and sample match, i.e. $n_{\mathrm{SIL}} = n_{\mathrm{sample}}$. Practical solid immersion lenses may work close to this condition rather than strictly meeting it.

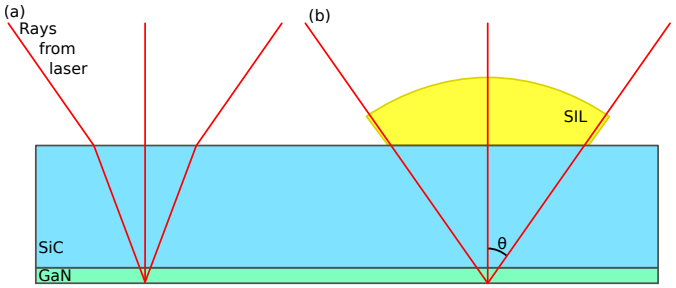


Figure 5.2: The working principle of a solid immersion lens (SIL): (a) focussing through a SiC substrate to the buffer-barrier interface, with no solid immersion lens; (b) with the addition of a solid immersion lens – no refraction occurs at the surface of the solid immersion lens or at the lens-substrate interface.

The spherical aberration introduced when focussing below the sample surface is removed by the solid immersion lens. In addition the working numerical aperture, defined as $NA = n\sin(\theta)$ for a refractive index n and a beam half-angle θ, is increased,[a] decreasing both the lateral and the axial diffraction limits (Equations 2.25 and 2.26). When measuring through the entire thickness of a wafer the benefit of a solid immersion lens will be greatest, as the spherical aberration increases with depth in the sample, while the benefit of a solid immersion lens is rather limited for imaging a few microns below the sample surface. This limited benefit combined with the difficulties of using a solid immersion lens in the presence of electrical contacts on the surface means that solid immersion lenses are generally only considered for imaging from the back of a wafer. A solid immersion lens can therefore be regarded as complementary to a spatial filter for these reasons, and the two may be used together to provide a further improvement to the axial resolution.

The use of a solid immersion lens requires excellent optical contact between the solid immersion lens and the sample: even an air gap of $^1/_{80}$ of the working wavelength degrades performance significantly.[166] This is a significant limitation if a solid immersion lens is to be placed on a hard crystalline material such as a SiC wafer – not only must a good match to the refractive index of SiC be achieved, but a perfectly smooth and flat surface be formed.

Alternatively a solid immersion lens may be formed directly into the back of the wafer, by photolithography[165] or the use of a focussed ion beam (FIB).[167,168] An example of an electroluminescence image collected using a solid immersion lens created with a FIB is shown in Figure 5.3. While a perfect optical contact to the rest of the substrate is assured by these methods, obtaining an accurate surface figure on the solid immersion lens is not simple, and the position of the solid immersion lens on the sample is fixed,

[a] As θ does not decrease when the beam enters the sample.

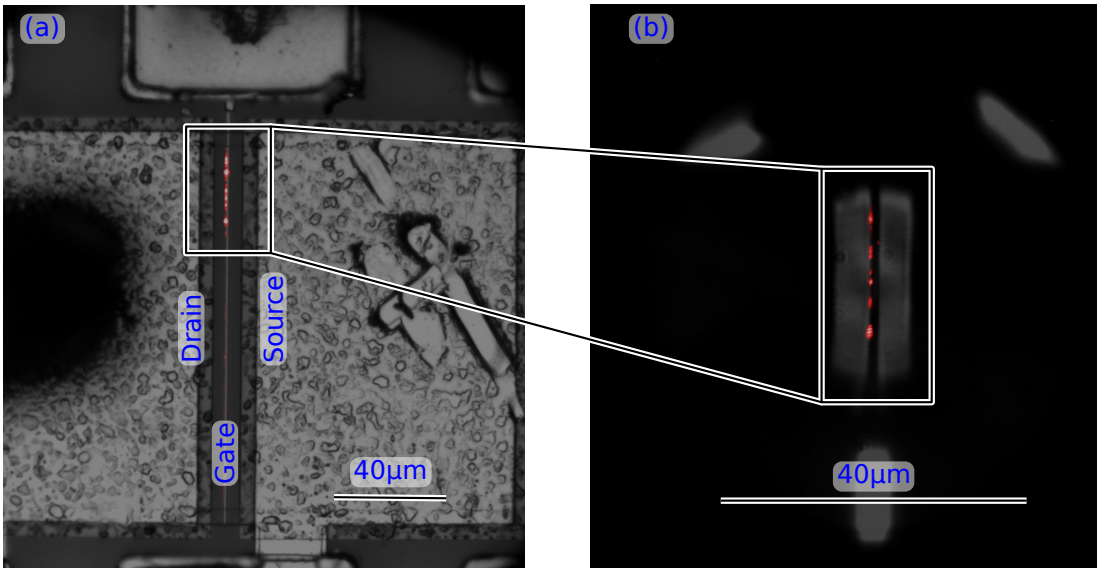


Figure 5.3: Electroluminescence images of a gated transfer length method test device: (a) from the top of the wafer, no solid immersion lens; (b) from the back of the wafer, with solid immersion lens. Images courtesy of J. Pomeroy (private communication).

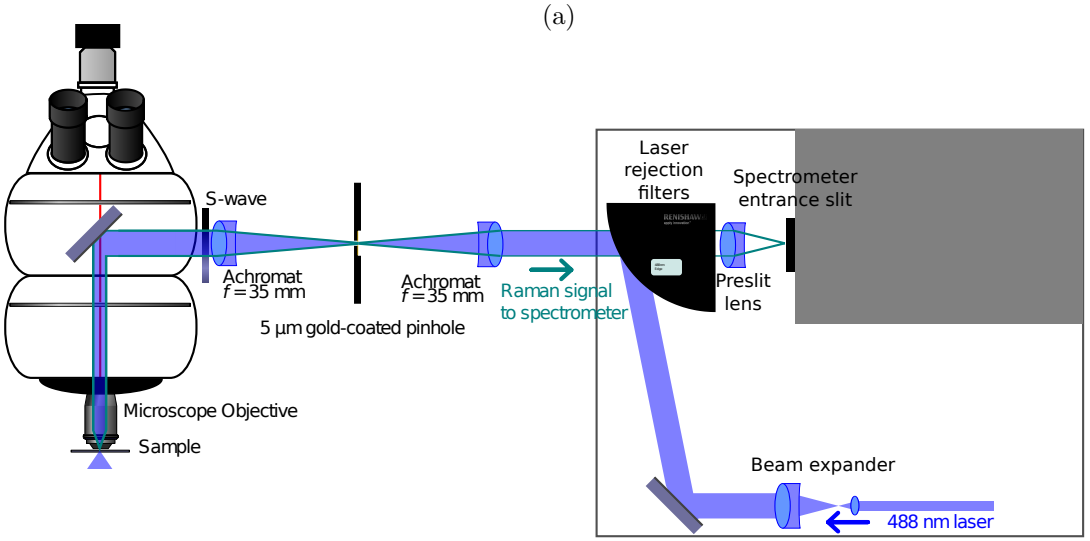
restricting the information that can be obtained to a small area, as may be seen from the curvature of the gate region in Figure 5.3(b). A solid immersion lens, while useful, can therefore be seen to not provide the solution to all the problems of depth resolution, and a more general solution is required, such as a spatial filter. Results obtained with a solid immersion lens are compared to those obtained using the spatial filter in Figure 5.5. A glass solid immersion lens was also used in the attempt to generate azimuthal and radial polarisations by focussing through calcite, which is discussed in Section 5.1.6

5.1.3 The design of the spatial filter

The spatial filter illustrated in Figure 5.4(a) was designed to have no overall magnification, meaning a pair of lenses of identical focal length. The upper limit on the total spatial filter length was determined by the available space, while the lower limit was set by the need to incorporate suitable opto-mechanical components to allow the components of the spatial filter to be aligned. These constraints may be seen in Figure 5.4(b). The focal length of each lens was therefore chosen to be 35 mm, with a pinhole size of 5 µm, approximately halving the sample-space pinhole size PH compared to the standard configuration.

The use of a pinhole (rather than the slit of the standard Renishaw system) and the reduced effective pinhole size inevitably caused a significant reduction in signal levels

(a)



(b)

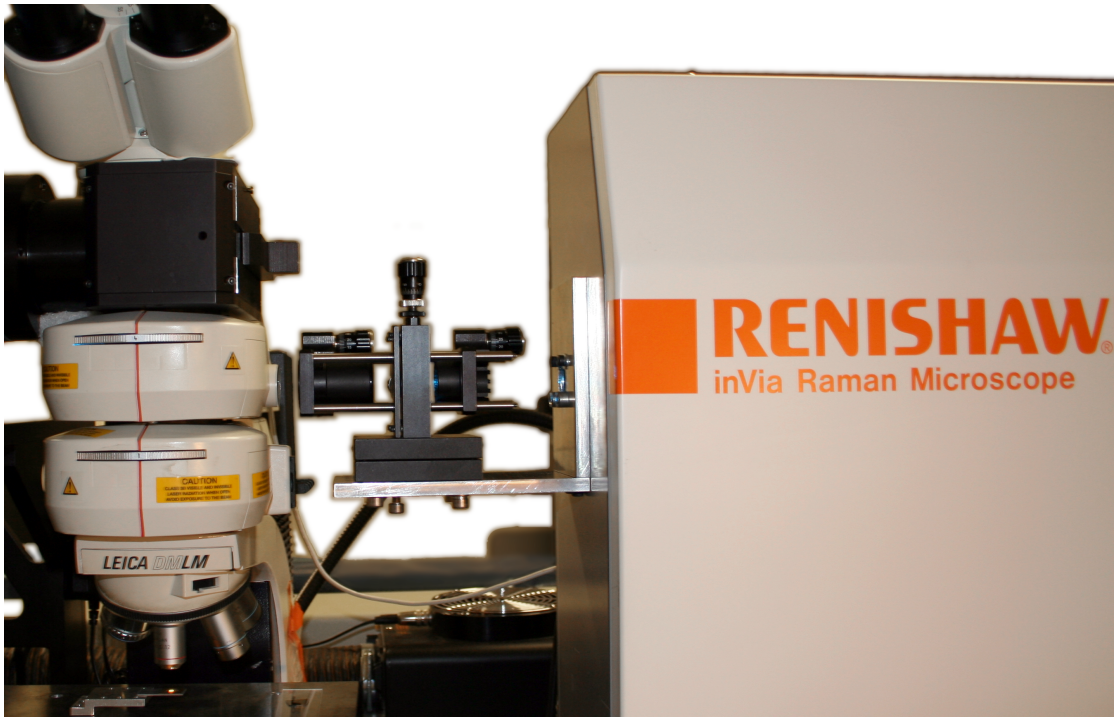


Figure 5.4: The layout of the spatial filter with respect to the sample and spectrometer: (a) schematic; (b) photograph (with the s-wave not present).

compared to a standard micro-Raman system. As some of the loss of signal was caused by a reduction in the laser power arriving at the sample due to losses at the pinhole, the power at the laser source was increased to provide comparable or slightly higher power levels at the sample to standard measurements, however signal intensities at the CCD were still ~5 % of those without the filter.

Because the pinhole was at a focus of the laser and imaged onto the spectrometer entrance slit, any fluorescence or Raman-scattered light from the pinhole foil was focussed onto the the entrance slit of the spectrometer and thus contributed to the background recorded at the CCD. The use of a gold coated pinhole was necessary to remove much of this background, which could be tentatively attributed to compounds in the stainless steel of a standard pinhole, and the associated noise. Even with a gold-coated pinhole the spectral background due to the spatial filter was not completely removed, which adversely affected the signal-to-noise ratio that could be obtained. Mechanical and thermal system drift limited the total measurement time (for all spectra within a depth map) to no more than 1 to 2 hours, which combined with the low signal strength meant that the E_2 phonon mode of the GaN was used despite its greater sensitivity to strain and smaller temperature dependence than the weaker A_1 mode, which could not always be located with sufficient precision for thermography.

The spatial filter was designed to be used with a long working distance objective of as high a numerical aperture as possible subject to the need for sufficient working distance to contact the sample with DC probes. Measurements from either the front (contact side) of the wafer or from the back, through the substrate were envisaged, and considered in the design. Measuring operating devices from the back of the wafer relied on using a retro-reflector under the microscope stage, as described in Section 3.1.4.

The benefits of the spatial filter and solid immersion lens are illustrated in Figure 5.5, which compares the signal intensity as a function of depth for the standard Raman system, the system with the spatial filter added, and the spatial filter combined with a solid immersion lens made from ZnSe, which has a refractive index of 2.66 at 532 nm, similar to that of SiC but in a softer material. These measurements were made on the surface of a silicon wafer, which is a standard test of system confocality, as in the ideal case there would be no signal for a focus either above or below the surface of the Si. A 0.75 numerical aperture objective lens was used. The improvement due to the SIL is less than would be expected from theory,[163] probably due to surface irregularities on

the solid immersion lens leading to an air gap between the solid immersion lens and the sample.[166]

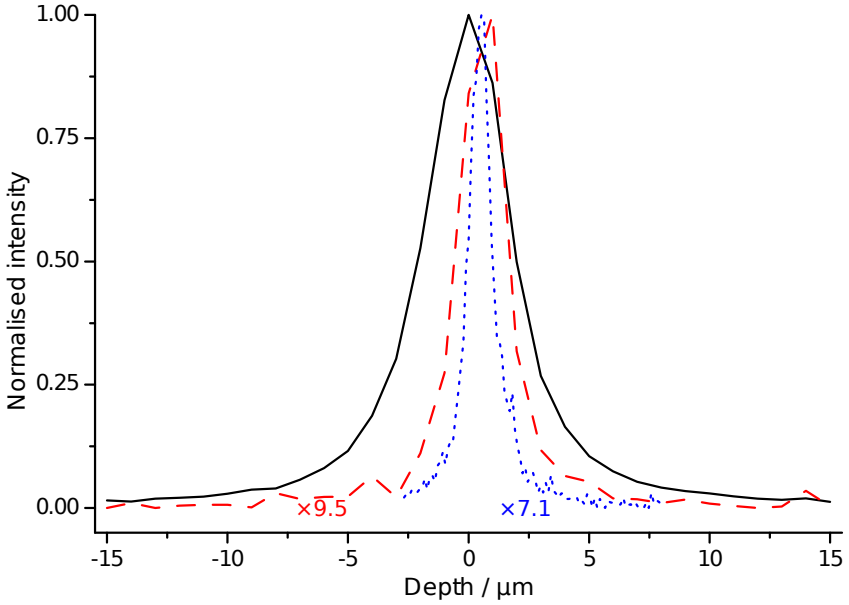


Figure 5.5: Raman depth series measurements at 488 nm on Si: standard system in high confocality mode (black curve, full width at half maximum 4.07 µm); with spatial filter (dashed red curve, full width at half maximum 1.81 µm); with spatial filter and ZnSe solid immersion lens (blue dotted curve, full width at half maximum 1.07 µm).

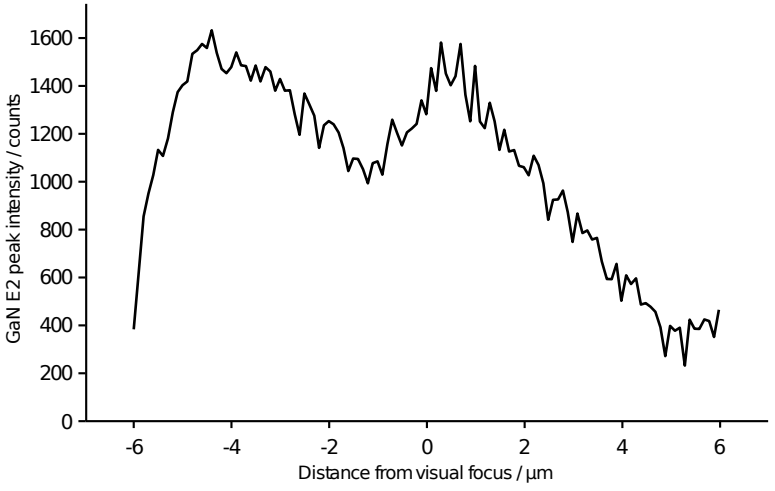


Figure 5.6: Raman depth profile measured from the back of a GaN-on-SiC wafer using a retro-reflector and an objective corrected for spherical aberration. The two maxima correspond to s- and p-polarised incident light.

As the spatial filter was intended for measuring samples from either the top or the bottom of the wafer, the effects of birefringence discussed in Section 2.3.4 had to be considered. The two foci produced by the s- and p-polarisations were first seen to be an issue when measuring from the back of a wafer. Variable coverslip correction was used in this case to reduce the adverse effects of refraction at the sample, but it can have no effect on the problems due to birefringence. Two intensity maxima in the E_2 Raman

peak were observed during depth mapping as shown in Figure 5.6, as well as two visual foci. For Δ of a few hundred microns in SiC, the maxima of the s- and p-polarisations are around 3 µm to 10 µm apart, which is enough to observe two separate focal planes. This makes obtaining independent spectra from the top and bottom of a typical GaN buffer impossible even after correcting for isotropic refractive effects, therefore optical systems for producing a beam where all rays are as shown in Figure 2.19(c) were investigated, i.e. azimuthal polarisation. Even when measuring from the top of the wafer the axial focus size is extended by the splitting of the focus into s- and p-polarised components, though this effect is insignificant over the thickness of the GaN buffer in a typical device.

5.1.4 Laser intensity distribution in a high electron mobility transistor

The laser intensity profile near a focus in a sample may be calculated from Equations 2.26, 2.28 and 2.33. These equations may be applied to a typical AlGaN/GaN HEMT to model the resolution that may be obtained in a thermography experiment. The following assumes a HEMT structure consisting of 1.9 µm of GaN on SiC. The nucleation layer and barrier ($d_{\mathrm{NL}} \sim 50$ nm of AlN and $d_{\mathrm{barrier}} = 25$ nm of AlGaN respectively) are ignored in optical calculations as their thicknesses d_{NL} and d_{barrier} are much less than the wavelength λ of the lasers used, and their refractive indices are close to that of GaN. Refraction at the GaN-SiC interface is also ignored (see Section 2.3.2) despite the SiC being thicker and having a refractive index n_{SiC} of 2.55. The resulting laser intensity distribution is shown in Figure 5.7.

Considering the laser intensity in the GaN (blue shaded region of Figure 5.7) it may be seen that a signal weighted towards the bottom of the GaN may be obtained by acquiring spectra at $\Delta = 2$ µm, which gives a spot centre at 5.7 µm below the surface, as only that part of the tail of the Gaussian distribution which extends into the GaN will contribute to the Raman peaks from the GaN. Similarly a signal weighted towards the top of the GaN may be obtained with $\Delta = -1$ µm, i.e. with a spot centred 1 µm above the surface. It is clear from Figure 5.7 that the signal will be greatly reduced compared to a focus at the sample surface, the conventional focus, as in that case nearly $^1/_2$ the laser distribution curve lies in the GaN, with some weighting towards the surface. This loss of signal is in addition to the loss due to the low throughput of the spatial filter, however by combining increased laser power with longer acquisition times, and using the E_2 mode

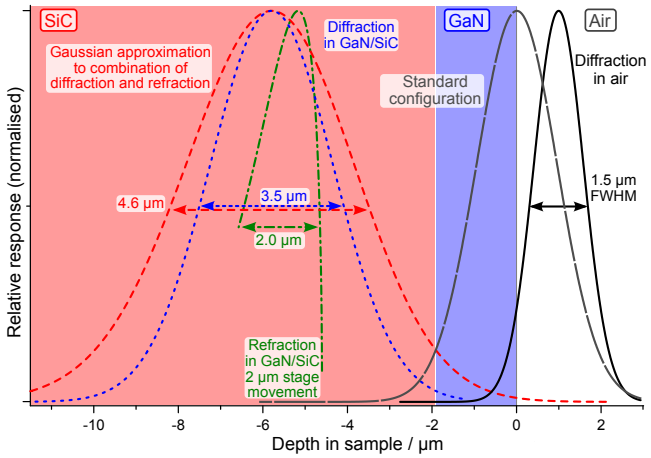


Figure 5.7: Laser intensity as function of depth in the sample, determined from modelling, with the laser above the sample surface (black solid line), and, for a sample movement of $\Delta = 2$ µm from a focus at the sample surface to inside the material, considering only diffraction (blue dotted line) and only refraction (green dash-dot line). The red dashed curve shows the combined Gaussian response approximation used in analysing simulated data. The wavelength used was 488 nm, the objective numerical aperture was 0.75 (focal length $f = 2.62$ mm), the sample refractive index was $n = 2.3$, and the confocal pinhole was 5 µm. The grey broken curve shows the equivalent distribution for the standard configuration of the Renishaw micro-Raman system in high confocality mode with the laser focussed on the surface of the device.

instead of the weaker A_1 (LO), sufficient signal may still be obtained to measure the temperature of the GaN. In a HEMT using the E_2 mode is feasible, as reference spectra in the pinched-off state allow the effect of strain on the peak position to be compensated (Section 3.2). It was important to ensure that these pinched-off reference spectra were taken under the same conditions of confocality and at the same depths as the on-state spectra.

5.1.5 Azimuthal and radial polarisations

Radial and azimuthal polarisations describe beams where the polarisation is parallel or perpendicular to the radius of the beam respectively, at all points in the beam, as will be shown in Section 5.1.7. A common feature of these two polarisation states is that the polarisation is undefined on the axis of the beam, where the intensity goes to zero. Azimuthal polarisation is desirable for avoiding the effect of birefringence when focussing along the optic axis of an optically anisotropic crystal, as was mentioned in Section 2.3.5. Radial polarisation has several uses of interest in Raman spectroscopy: the lateral dimension of the focus of a radially polarised beam can be smaller than for a linearly polarised beam;[169] in tip enhanced Raman spectroscopy radial polarisation increases the part of the signal due to the tip;[170,171] and, of most interest here, a radially

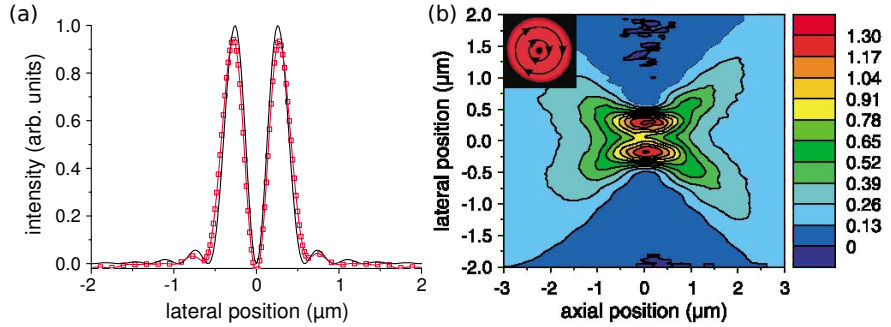


Figure 5.8: The focus of an azimuthally polarised beam: (a) measured (squares) and calculated (lines) intensity distribution in the focal plane; (b) measured intensity as the beam is defocussed, the inset shows the polarisation within the beam. Figure taken from Reference 87.

polarised beam may be used to give access to phonon modes in orientations where they are not normally available, as will be shown in Section 5.1.6. Both the usefulness of radial polarisation in tip enhanced Raman spectroscopy and the access to additional phonon modes in a given geometry are due to the fact that a significant component of the light at the focus is polarised parallel to the bulk propagation direction of the beam – which corresponds to the field of a dipole parallel to the beam axis.[172] This is known as z-polarisation.

Azimuthal polarisation is not particularly well studied compared to radial polarisation, but the intensity distribution at the focus has been measured and modelled.[87] The focus is annular as shown in Figure 5.8. This annular focus may degrade the lateral resolution compared to conventional linear or circular polarisation but the effect on the axial resolution is not so simple. It is clear from applying Equation 2.35 to Equation 2.28 (Section 2.3.4) that the p-polarisation at all points in the beam resulting from azimuthal polarisation will have a smaller axial spot size after focussing through a thick SiC substrate, but as the separation of the s- and p-centroids is linear in Δ the effect at or near the surface of the sample is much weaker: $\sim$50 nm of separation at $\Delta = 2$ µm. The additional loss of focus due to birefringence can therefore be neglected except when measuring from the back of the substrate with an objective corrected for spherical aberration.

Azimuthal polarisation still improves the contrast between the measured peak positions at the top and bottom of the GaN buffer in experiments far from metal contacts. The reasons for this are not completely clear, but as may be seen in Figures 5.8 and 5.10 the intensity on the axis of the beam is low or zero compared to the surrounding regions, even if nothing is done to block the centre of the beam. This annular beam structure

is observed however the azimuthal polarisation is produced, as the intensity must go to zero on axis, where the polarisation is undefined,[173] and is true both at the focus and in collimated light. In general reducing the on-axis intensity component of a beam has some benefit in reducing the axial spot size; in the extreme case a ring-like apodising filter may be used.[174] In addition the limited overlap of the laser focus with the conventionally-shaped focal region which can be observed by the spectrometer is likely to suppress Raman-scattered light generated away from the target focal plane.

5.1.6 Azimuthal polarisation by focussing through a calcite plate

The same effect that leads to two foci when focussing deep inside a SiC wafer due to the birefringence of the SiC can be used to provide two axially-separated foci in air. One of these foci corresponds to radial polarisation, the other to azimuthal, therefore a pinhole may be used to select the polarisation of interest.[85]

As the birefringence of SiC is rather low ($\delta n_{\mathrm{SiC}} = n_{\mathrm{e}} - n_{\mathrm{o}} = 0.043$ at 532 nm) the separation of the foci for thin SiC wafers is too small to allow the selection described above. Calcite ($CaCO_3$) is well known for its large birefringence ($\delta n_{\mathrm{CaCO_3}} = 0.175$ at 532 nm) and is commonly used in the manufacture of optical components, including for generating radial and azimuthal polarisation,[84] therefore a calcite plate 1 mm thick was polished and placed in the laser beam between two objective lenses before the system beam expander. Circularly polarised incident light was used. To minimise the effects of spherical aberration on entry into the calcite plate a glass solid immersion lens was bonded to the calcite plate, using an index-matching optical adhesive. The solid immersion lens was in the form of a truncated hemisphere such that the total thickness of the calcite and solid immersion lens was that of a hemisphere of radius 2 mm. Blocking the centre of the beam using a mask before the calcite plate was used to reduce the unwanted polarisation component in some experiments. This is because the indices of refraction experienced by the radial and azimuthal components are closer near the axis, as may be seen from Equation 2.35. This central block or mask, when used with radial polarisation, enhances the z-component of the polarisation at the sample,[86,175] i.e. the polarisation in the propagation direction. This system is set out in Figure 5.9.

Despite the tendency of calcite to cleave tetrahedral pits from the surface when polished normal to the optic axis,[176] azimuthal (Figure 5.10) and radial polarisations were

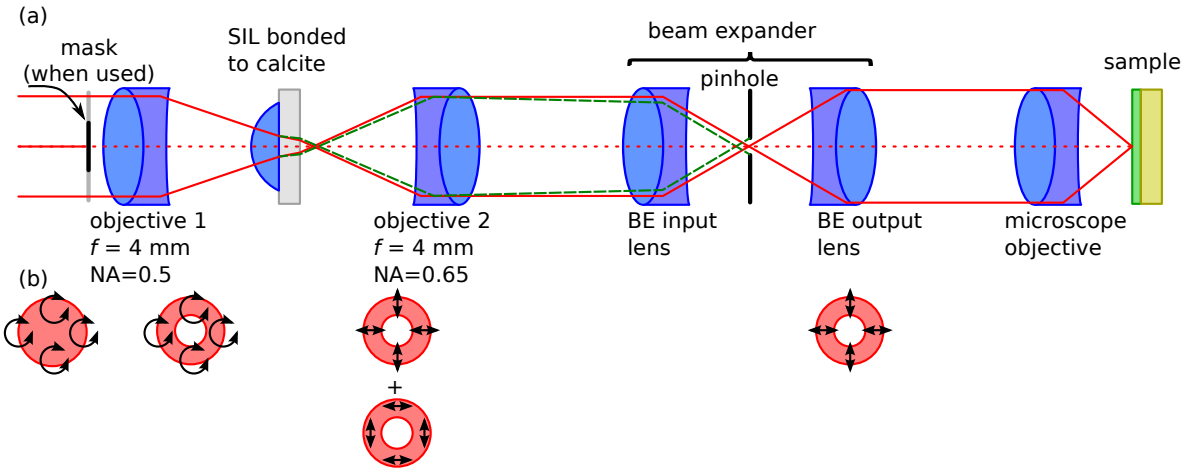


Figure 5.9: Using a calcite plate to obtain a radially polarised beam and therefore a component of z-polarisation at the sample: (a) Ray diagram; (b) Beam cross-section and polarisation at selected points along the beam path. If objective 2 is refocussed, azimuthally polarised light (green dashed line) will pass through the pinhole.

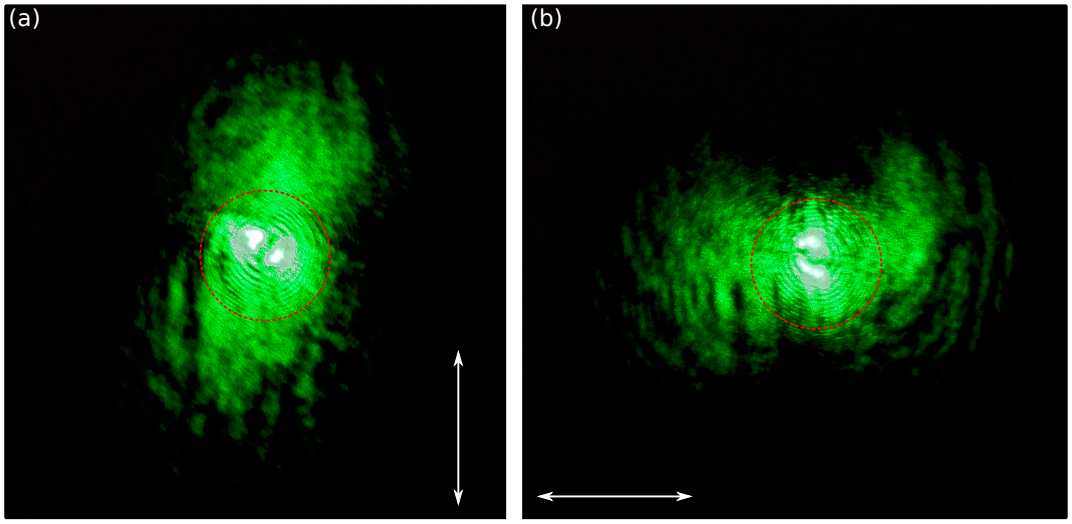


Figure 5.10: The polarisation of the laser after passing through a beam expander containing a calcite plate to separate radial and azimuthal polarisations. The collimated beam was passed through (a) a vertical linear analyser; (b) a horizontal linear analyser. The red dotted circle indicates the approximate size of the beam itself, light outside this circle is scattered at the surfaces of the calcite. No mask (Figure 5.9) was used.

obtained. The beam quality however was poor as the tetrahedra cleaved from the surface caused scratches in polishing, which then caused light scattering. This is evident in the light outside the real beam size (indicated in Figure 5.10), as well as in the non-zero intensity at the centre of the beam. The scattering was mainly on the output face of the calcite plate as the index-matching adhesive used to attach the solid immersion lens to the input face filled scratches on that face, although not perfectly as the refractive indices are not a perfect match i.e. $n_{\text{glass}} \approx n_{\text{adhesive}} \approx \bar{n}_{\text{CaCO}_3}$, where $\bar{n}_{\text{CaCO}_3} = {}^1\!/_2(n_{\text{e-CaCO}_3} + n_{\text{o-CaCO}_3})$.

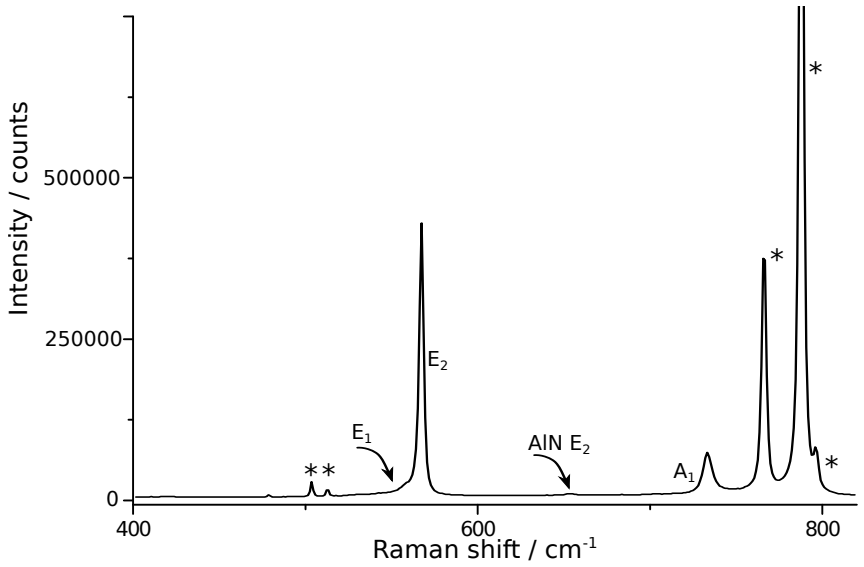


Figure 5.11: Raman spectrum in radial polarisation of a GaN epilayer on SiC at 532 nm. Peaks marked $*$ are due to the SiC substrate according to References 177–180.

Radial polarisation obtained using the calcite plate allowed the E_1 mode normally invisible in the $z(x-)\bar{z}$ polarisation to be observed as a shoulder on the E_2 peak. The measuring geometry at the sample may then be described as a mixture of $z(x-)\bar{z}$ and $z(z-)\bar{z}$. Curve fitting found the E_1 peak height to be 2.5% of the E_2 peak height. Similarly, additional peaks from the SiC substrate were observed due to the polarisation as well as due to the long acquisition time, which enabled the E_2 mode of the AlN nucleation layer to be observed. All these peaks are shown in Figure 5.11.

The E_1 mode would be useful for separating the in-plane and normal strain components,[179] but it would need to be clearly resolved from the E_2 mode for the E_1 peak position to be measured with sufficient accuracy for this. A much larger z-polarisation component with respect to the in-plane component would therefore be required. Using the s-wave polarisation converter discussed in the next section combined with an optimised mask would go some way towards achieving this. A much higher numerical aperture would also be required, for example by using an oil-immersion objective, which is only possible if electrical contact to the sample is not needed.

5.1.7 The s-wave polarisation converter

When it proved impossible to obtain sufficient beam quality by using the calcite plate beamsplitter (as may be seen from Figure 5.10) the concept of a spatially-varying waveplate was investigated.[86,87] These typically use a small number of waveplate segments

(most commonly four or eight) cut at appropriate angles to their optic axes and assembled to rotate the polarisation of segment of the beam by differing amounts, and are usually made specifically for a particular experiment taking into account the beam size and wavelength. A segmented waveplate may be used to achieve radial or azimuthal polarisation, by rotating it to the correct angle relative to the polarisation of the incoming light.

A recent development, equivalent to a segmented waveplate with a much larger number of segments, is the use of a femtosecond laser to create nanostructures in silica.[88] This component is commercially available under the name s-wave, and is shown in Figure 5.12.

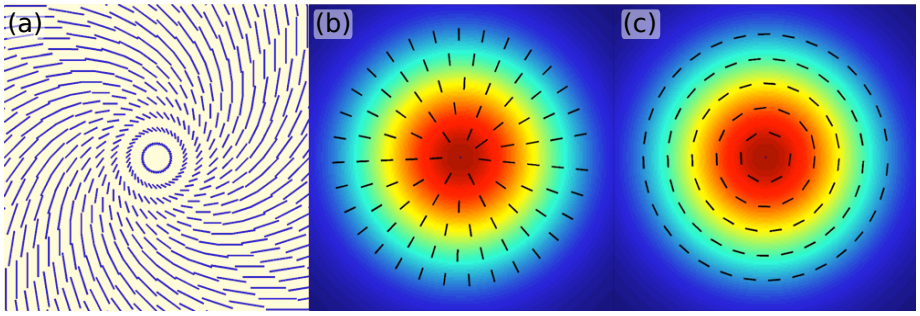


Figure 5.12: The s-wave polarisation converter: (a) schematic drawing of nanograting distribution; (b) distribution of the electric field for left circularly polarised light after the s-wave (radial polarisation); (c) distribution of the electric field for right circularly polarised light after the s-wave (azimuthal polarisation). Figure from Reference 88.

The s-wave was also more convenient to mount in the inVia micro-Raman system and to align than an additional spatial filter containing a calcite plate. It did however have an effect on the lateral distribution of light at the focus, which required it to be placed after the spatial filter, as shown in Figure 5.4(a) to reduce the attenuation of the laser at the pinhole it caused if placed before the spatial filter. Even in the position in which it was used the s-wave caused a significant loss of signal with respect to the spatial filter alone, by introducing a small amount of divergence into the beam.

5.2 Experimental and thermal simulation details

The experiments described in this section were all carried out at 488 nm with the spatial filter and azimuthal polarisation. The laser power at the sample, while higher than in typical thermography experiments, was less than 50 mW and constant within an experiment (both for on-state and pinched-off spectra), therefore laser heating of the

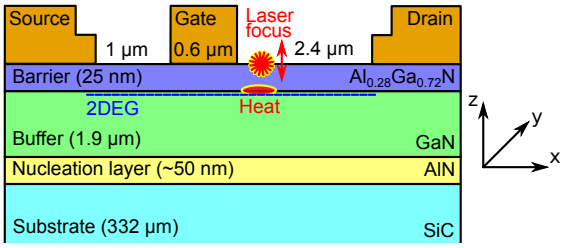


Figure 5.13: Schematic of the HEMT structure studied and simulated when measuring the temperature at the top and bottom of the GaN buffer.

samples may be neglected. An objective lens with a numerical aperture of 0.75 was used (Zeiss LD Plan-Neofluar 63×).

5.2.1 Measuring strain gradients in a GaN buffer

Initial measurements were taken on an unpowered transfer length method test structure, over 20 µm from the nearest metal contacts, ensuring that there was no possibility of reflections from contacts changing over a series of measurements and affecting results. The GaN buffer was 1.9 µm thick, on a 4-H SiC substrate, with a ~50 nm AlN nucleation layer. The barrier was 25 nm of $Al_{0.28}Ga_{0.72}N$. The Raman shift as a function of depth was measured by translating the sample between acquiring spectra, to produce a depth map. Nine depth maps were taken, the average peak position and intensity results of which are given in Section 5.3.1. By selecting data points which preferentially sample the top and bottom of the GaN buffer, the difference in strain (found from the difference in the E_2 peak position) between the GaN-SiC and AlGaN-GaN interfaces was measured.

5.2.2 Vertical temperature gradient in an operating high electron mobility transistor

A HEMT with the same layer structure as the transfer length method structure described in Section 5.2.1 was measured by Raman thermography depth mapping. Maps were acquired both in on state and pinched off, to allow the shift in peak position as a result of self heating to be extracted from the measurements despite the change in strain and therefore peak position with depth shown in Section 5.3.1. The source-drain spacing was 4 µm, with a 0.6 µm gate 2.4 µm from the drain, and the device width was 50 µm. The device structure used is shown in Figure 5.13.

The devices were operated under DC bias to 30 V in on state at zero gate-source bias; in pinched off-state a gate-source bias of −6 V was applied. The backplate temperature at the back of the SiC substrate was fixed at 25 °C. The depth maps were taken next to the gate, on the side nearest the drain, i.e. in the peak field region where the temperature may be expected to be at a maximum. The microscope stage was scanned along z, in increments of 0.5 µm or 1 µm, from a focus above the device surface into the GaN buffer layer and the substrate.

The signal close to a contact such as the gate was attenuated by some of the laser light and some of the Raman scattered light being lost to reflections at the gate. This necessitated longer acquisition times (60 s to 150 s) than when measuring far from contacts on transfer length method structures (30 s) in order to achieve sufficient signal-to-noise ratio for accurate peak positions to be obtained. Sample drift over the course of a measurement therefore became an issue, with the effect depending on the direction of the drift (Figure 5.13, as the device was aligned to the axes of the microscope stage). Drift in y could be ignored as the gate was aligned parallel to the y-axis. Drift in x could be detected only after the depth map had finished by examining the position of the device in white light. It therefore led to the complete loss of any map in which it occurred. Drift in the z-position of each depth map could be corrected by offsetting the z-scale on the assumption that the maximum Raman signal was recorded when the laser was focussed at the surface of the sample, an assumption confirmed by measurements including those shown in Figure 5.15. This, and variability in the initial move from a manual focus, accounted for the need for sub-micron z-steps.

5.2.3 Thermal simulation for comparison with depth-resolved thermography

As is common to extract the peak temperature from the experimentally-measured temperature,[158,181] the device temperature was simulated using *Thermal Analysis System (TAS)* finite difference software using thermal conductivities of the GaN (160 $\mathrm{W\,m^{-1}\,K^{-1}}$) and SiC (420 $\mathrm{W\,m^{-1}\,K^{-1}}$) similar to References 56 and 75. The process of thermally simulating a device is discussed in more detail in Section 3.5. The thermal boundary resistance between the GaN and the SiC was adjusted to fit the experimental data. This effective thermal boundary resistance contains the low thermal conductivity of the

AlN nucleation layer, commonly present between the GaN and the SiC substrate, and impacts from the very high defect density GaN just near the AlN nucleation layer.

To compare the simulated results properly to the experimental data it was necessary to post-process the simulation output. This is because the simulation calculated the temperature of nodes within the model. Whether these nodes were placed manually or automatically, they were at arbitrary locations, and were infinitesimal, and therefore poorly representative of the temperature which would be measured by a finite laser spot. Because of the variation in laser intensity within the spot a simple average temperature would not be suitable for comparison with experimental data. The uneven node distribution meant that an average weighted purely by laser intensity would be biased towards regions of high node density, for example the substrate temperature would be heavily biased towards the top of the substrate. As only a small region of interest within the model (i.e. a small subset of the nodes) corresponded to the Raman experiment, the subsequent analysis could take place on a much smaller number of nodes than were present in the model. The steps involved in converting the simulation output to a format suitable for comparison with the experimental data are shown in Figure 5.14, and the source code is given in Appendix A.2. Briefly the process consists of: reducing the *Thermal Analysis System* output to a manageable size and region of interest; forming a list of epilayers from user input; iterating over a specified range of points in three dimensions; and generating a weighted average temperature at each point from nodes within the focal volume around that point.

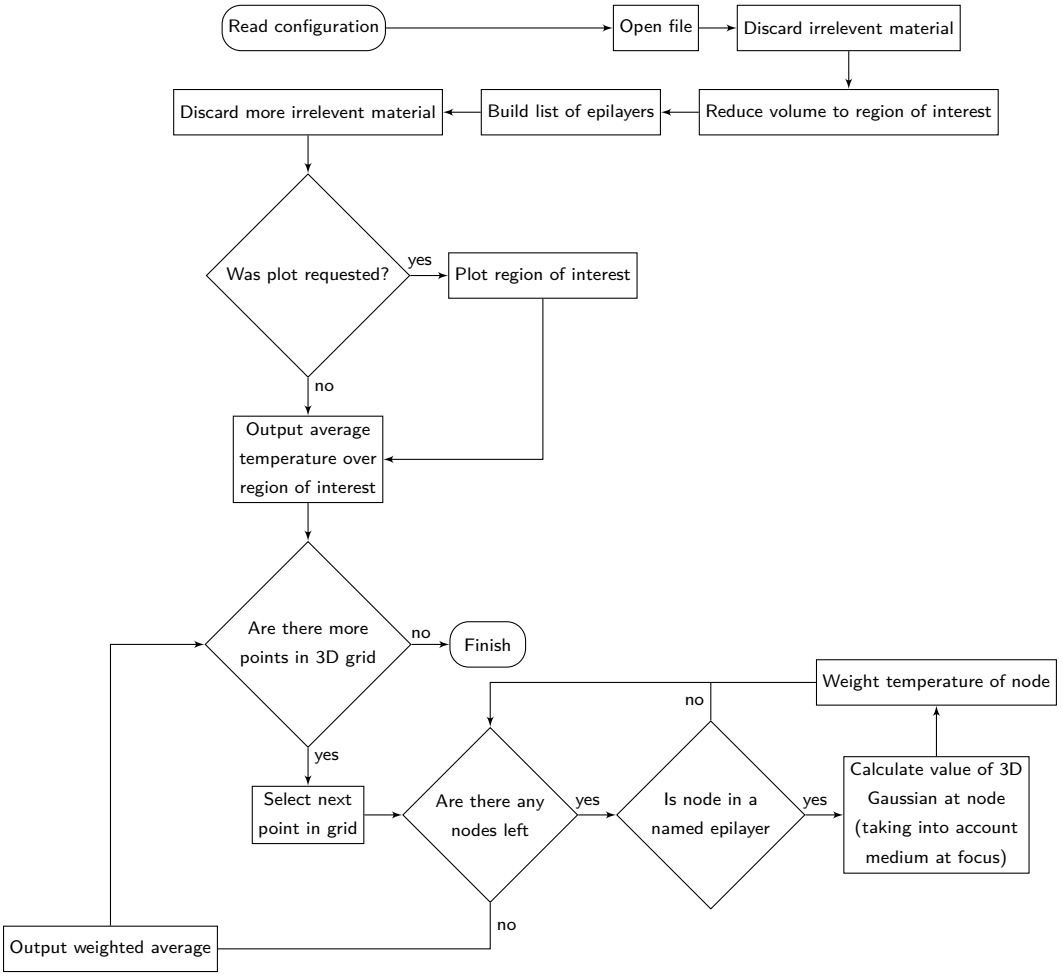


Figure 5.14: Simplified flow chart of the process of extracting temperatures averaged over physically realistic volumes from simulation data

5.3 Results and discussion

Some preliminary results have already been shown in Figures 5.5 and 5.11 to illustrate the methods of improving the spatial resolution and optimising the polarisation. The aim of these modifications is to show a temperature gradient between the top and bottom of the GaN buffer on an operating HEMT, however it is first necessary to ensure that these results will be free of artefacts by measuring the difference in strain through the thickness of a GaN buffer.

5.3.1 Strain in a GaN buffer

The E_2 peak position in an unpowered transfer length method as a function of depth is shown in Figure 5.15. Only changes in the peak position with depth, which correspond to changes in strain, are considered here. The E_2 phonon frequency is particularly sensitive

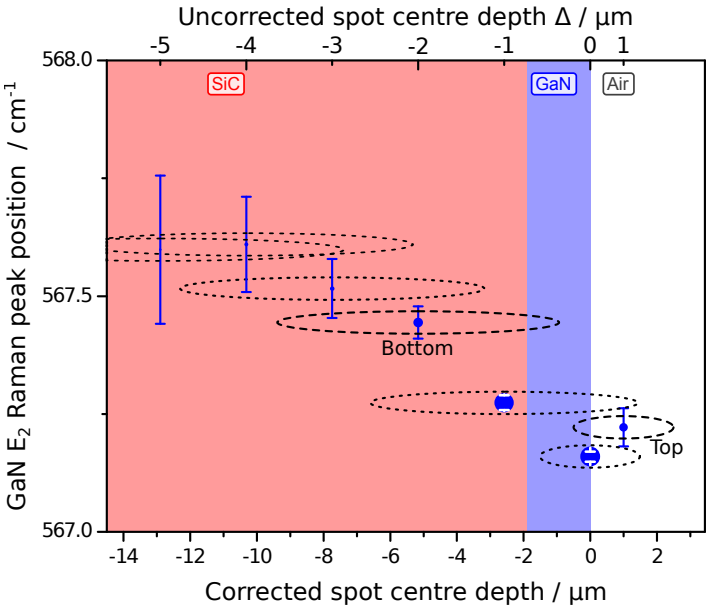


Figure 5.15: GaN E_2 Raman band position as a function of depth in a transfer length method test structure. The spot centre depth is corrected according to Equation 2.29, the dotted and dashed ellipses indicate the $^1/_{e^2}$ spot sizes found from Equations 2.33 and 2.25, and the symbol sizes are proportional to the Raman peak height. The difference in the E_2 peak position between the data points indicated by "Top" and "Bottom" and dashed ellipses corresponds to a difference in biaxial stress of $\sigma_{\mathrm{Bottom}} - \sigma_{\mathrm{Top}} = (83 \pm 22)$ MPa.[179]

to in-plane strain, which would be expected to change rapidly over the thickness of a GaN epilayer due to the lattice mismatch at the interface to the substrate (SiC in this case). A clear difference of (0.23 ± 0.06) cm^{-1} between the marked data points most closely corresponding to the top and bottom of the GaN layer is seen. Using the linear relationship between phonon frequency ω and stress σ, $\frac{\mathrm{d}\omega}{\mathrm{d}\sigma} = K^R_{\mathrm{E}_2^{\mathrm{high}}}$,[179] where for the E_2 (high) mode the linear factor $K^R_{\mathrm{E}_2^{\mathrm{high}}} = (2.7 \pm 0.3)$ cm^{-1} GPa^{-1} allows the difference in stress between the top and bottom of the GaN to be calculated. This difference is (83 ± 22) MPa taking into account the uncertainties in the peak positions and $K^R_{\mathrm{E}_2^{\mathrm{high}}}$. This means that the bottom of the GaN, i.e. at the GaN-SiC interface, is under compressive stress with respect to the top of the GaN layer. As the a lattice constant of SiC is less than that of both the GaN buffer and the AlN nucleation layer, this result has the expected sign. At the top and bottom of the GaN, the signal intensity, indicated by the sizes of the circles, was considerably less than for a focus at the sample surface, leading to an increased uncertainty in the peak position. However compared to deeper points where the $^1/_{e^2}$ spot size did not extend into the GaN buffer (i.e. only the tail of the laser distribution contributes to the measurement) a clear signal could still be obtained given an acquisition time of one minute per point.

The magnitude of the strain difference relies on a suitable value for $K^R_{\mathrm{E}_2^{\mathrm{high}}}$, for which

values in the literature vary, though the sign is consistent. Much of the variation is due to experimental differences and different substrates. Although measurements of the strain in GaN epilayers have been made using Raman spectroscopy in similar scattering geometries to this work, they tend to have been on sapphire substrates. On thin[182] GaN epilayers the polarisation was not specified,[a] while measurements in $z(yy)\bar{z}$ configuration were on thick[183] GaN epilayers. It is not clear whether corrections for the centre of the focal spot were made in either of these cases. For GaN layers grown on 6H-SiC the phonon mode shift with strain has been studied in some detail. Raman spectroscopy was used, in a different geometry to this work, as well as x-ray crystallography, allowing the relationship between strain and peak position to be found for the E_2 (high) peak used here among others.[179] The scattering geometry used for measurements of the E_2 peak position was $x(yy)\bar{x}$ in this case at an edge of the sample. This approach allows the lateral rather than the axial resolution to determine the size of the measurement area, giving finer resolution than could be obtained here and avoiding the need to correct for refraction. This sampling geometry is not compatible with studies of operating or indeed operable devices as it requires optical access to the edge of the wafer at the measurement point, and a cut through a device, even if it did not destroy the device, could be expected to change its characteristics. It is also not possible to rule out the possibility of the strain being affected by measuring at a cut and polished edge rather than in the middle of a wafer.

Direct comparisons of the magnitude of the change in stress to previous work are complicated by the fact that studies of stress as a function of depth are generally carried out on much thicker epilayers, or by comparing samples with different epilayer thicknesses. The results found here do however appear to broadly agree with the work of Röder *et al.*,[183] whose results show a steeper change in peak position over thinner layers when using confocal Raman, however their thinnest layer was 25 µm and was grown on sapphire with strain relief layers. As well as the dependence of strain on the parameters of the nucleation layer, the change in strain across the buffer is interdependent with the threading dislocation density in the buffer,[184] as dislocations provide a mechanism for the relaxation of the stress which causes them. These results confirm the sensitivity of the method to Raman shift and therefore provide a sound basis for thermography studies.

[a] But it may be assumed to be $z(y-)\bar{z}$ from other details given.

5.3.2 Difference in temperature across the GaN buffer in AlGaN/GaN HEMTs

Figure 5.16(b) shows the measured temperatures as a function of depth 0.6 µm from the drain side of the gate contact of a HEMT, where peak heat dissipation occurs,[75] from the top GaN surface into the SiC substrate. To enable measurement of temperature of the top and the bottom of the GaN layer from the curves in Figure 5.7, spectra weighted towards the top and bottom of the GaN buffer were obtained by offsetting the sample surface from the focus as formed at the sample surface, by 1 µm into the air, or 2 µm into the sample, which correspond to the two data points indicated. Compared to standard Raman thermography, therefore not only the average temperature of the GaN layer, but the temperature gradient inside the GaN can be directly measured. Also apparent in Figure 5.16(a) is the temperature drop from the bottom of the GaN buffer layer in the AlGaN/GaN HEMT to the SiC illustrating the presence of an effective thermal boundary resistance at this interface.

The simulation which was fitted to the experimental data is illustrated in Figure 5.16(a). The good agreement between the experimental data and the simulation, which used a homogeneous GaN thermal conductivity, is a clear indication that it is reasonable to treat GaN as having a single thermal conductivity throughout its depth in a thermal device simulation. This is despite the gradient in defect density through the GaN layer thickness. Although there are only two temperature data points in the GaN and one in the SiC, it is still possible to test whether it is possible to treat the GaN as having a single thermal conductivity. If a high defect density region having a low thermal conductivity[185] extended far into the GaN, the "top" GaN temperature would increase and the "bottom" would decrease with respect to the simulation shown in Figure 5.16. The simulation would then not be able to fit this without a low GaN thermal conductivity and a low thermal boundary resistance, both of which would disagree with previously reported values for comparable samples.

A thermal boundary resistance across the GaN-SiC interface of 1×10^{-8} m^2 K W^{-1} in the simulation resulted in good agreement with the experimental data. This contains contributions from phonon mismatch between the GaN buffer and the AlN nucleation layer, and between the nucleation layer and SiC substrate, however the main contributions arise from the low thermal conductivity of the high defect density AlN and the high defect density GaN very near to this interface.[55,56,186] This is limited to a very narrow

region around the GaN-AlN interface consistent with transmission electron microscopy data[55] making it possible to treat the remainder of the GaN buffer layer as material of homogeneous thermal properties to a first approximation, as discussed above. This layer would then be too thin to have any significant effect on the GaN Raman peak positions, especially as its low crystal quality would tend to broaden the Raman peaks from it, thus spreading its contribution to the spectrum. The value for the thermal boundary resistance found here is consistent with the value reported in Reference 55.

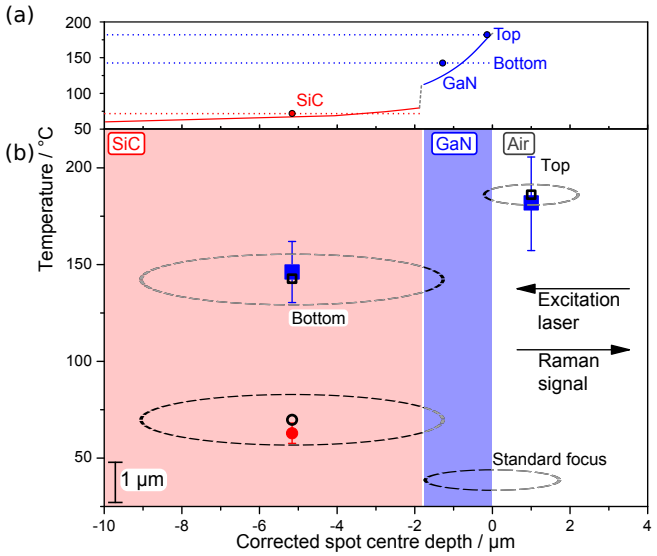


Figure 5.16: (a) Simulated temperature (solid lines) at the drain side of the gate contact of an AlGaN/GaN HEMT, i.e. at the device temperature peak, as a function of depth from the device surface through the GaN buffer layer and the SiC substrate. The dotted lines are the measured temperatures of the top and bottom of the GaN, and of the SiC, indicated by points at depths considering the vertical averaging of temperature in the measurement due to the laser intensity profile. (b) Measured temperatures of the GaN (blue squares) and SiC (red circle) with corrected (Equation 2.29) focus depth. Open black symbols indicate temperatures extracted from simulations using physically realistic weighted averaging, dashed ellipses indicate the $^{1}/_{e^2}$ size of the Gaussian volume (Equations 2.25 and 2.33) over which the measurement averages temperature and the simulation data was averaged to compare to the experimental data.

Obviously also for the measured data points volume averaging of temperature needs to be considered which was taken into account in Figure 5.16(a), and is discussed in the following in more detail using Figure 5.16(b). Two experimental data points in Figure 5.16(a) are considered, one from near the top surface and one from near the bottom surface of the GaN buffer layer, marked in Figure 5.16 by “Top”and “Bottom”. Indicated in Figure 5.16(b) is the extent of the focal volume for these points determined from Equation 2.33 illustrating that the bottom and top parts of the GaN buffer layer are preferentially probed in this measurement. When comparing to the simulated thermal device data, the simulated nodal temperatures in the GaN layer and SiC substrate,

shown in Figure 5.16(a), were therefore averaged with a weight considering the three-dimensional distribution of the light intensity. An example of this process is shown in Figure 5.17.

The light intensity in the laser spot is given by the three-dimensional Gaussian distribution

$$I(\mathbf{r}') = I(x', y', z') = \frac{1}{\sqrt{2\pi^3\sigma_x^2\sigma_y^2\sigma_z^2}} \exp\left[-\left({}^{x'}/_{\sigma_x}\right) - \left({}^{y'}/_{\sigma_y}\right) - \left({}^{z'}/_{\sigma_z}\right)\right] \tag{5.1}$$

where $\sigma_x = \sigma_y = {}^{w_{\text{lat}}}/_4$ and $\sigma_z = {}^{w_{\text{ax}}}/_4$, as the ${}^1/_{\text{e}^2}$ widths w_{lat} and w_{ax} calculated from Equations 2.25 and 2.33 are by definition equal to 4σ. The axial spot size w_{ax}, and therefore σ_z, change with the depth Δ, while the lateral spot size w_{lat} undergoes a step-change at $\Delta = 0$, the sample surface. The coordinates x', y' and z', i.e. $\mathbf{r}'$ take as their origin the focus depth calculated from Equation 2.29, and hence also include a dependence on Δ. To find the temperature equivalent to that measured at position $\mathbf{r} = (x, y, z)$, denoted by $\bar{T}$ to indicate that this is spatially averaged, it is necessary to numerically evaluate

$$\bar{T}(\mathbf{r}) = \frac{\int T(\mathbf{r}-\mathbf{r}')V(\mathbf{r}-\mathbf{r}')I(\mathbf{r}')\,\mathrm{d}\mathbf{r}'}{V(\mathbf{r}-\mathbf{r}')\int I(\mathbf{r}')\,\mathrm{d}\mathbf{r}'} \tag{5.2}$$

where $T(\mathbf{r}-\mathbf{r}')$ is the temperature of a node at absolute position $\mathbf{r}-\mathbf{r}'$ from the simulation, and $V(\mathbf{r}-\mathbf{r}')$ is the volume corresponding to that node. It was necessary to take into account the volume surrounding each node as the nodes were not generally evenly spaced, as may be seen in the SiC substrate in Figure 5.17. This was done by generating a Voronoi tessellation seeded with the locations of the nodes; the volume of each node was then taken to be the volume of the Voronoi cell in which it lay. Typically in the buffer a regular grid was used (as shown in Figure 5.17), but in the substrate the grid was automatically generated and the spacing between the nodes (and therefore the volume surrounding each node) varied significantly. This allowed closely-spaced nodes in regions of high thermal gradients (i.e. near the heat source where it is important to be able to obtain the peak temperature), and much larger spacing far from the heater where the temperature changes only slowly, minimising the number of nodes needed to obtain an experimentally-relevant simulation. Although the limits on the integration would ideally extend to $\pm\infty$, or at least to the edges of the model, this is not necessary in practice. The example shown in Figure 5.17 integrates from $-8\sigma_i$ to $8\sigma_i$, where i is

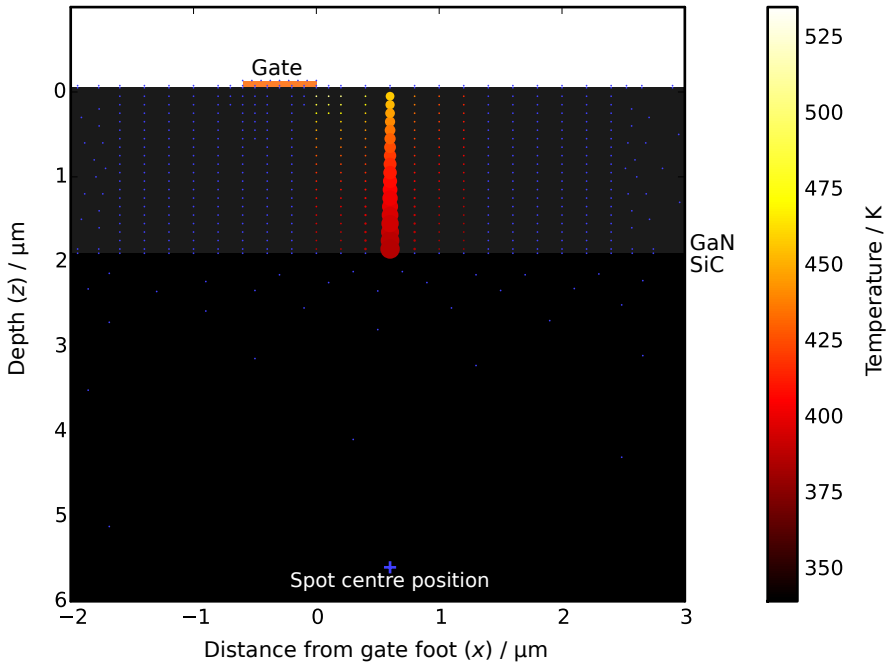


Figure 5.17: Calculating the volume-averaged temperature within the GaN buffer of an AlGaN/GaN HEMT for a laser spot centred at $\Delta = 2$ µm (blue +). The area of the data points, which are the nodes of a thermal simulation, indicates their weight in the average. The colour of the points indicates their temperature, with blue points indicating the positions of nodes excluded from the average as they are far from the spot centre or not in the GaN.

one of x, y, and z; unless this would extend beyond a layer boundary (in z) or the edges of the model.

The distinct difference in temperature between the top and the bottom of the GaN is due to the heat generated near the gate contact at the GaN top surface, or more exactly at the interface between the GaN buffer and AlGaN barrier, as illustrated in Figure 5.13, with heat extracted from the device through the GaN-SiC interface and the SiC substrate to a heat sink at the back of the SiC substrate. The temperature gradient inside the GaN layer can therefore clearly be also measured directly rather than just taken from a thermal simulation, also providing a further handle to refine thermal simulations to improve their accuracy. These results validate the approach taken of improving the depth resolution of Raman thermography using non-contact optics from the top of an operating device.

5.4 Conclusions

Methods of improving the depth resolution of Raman thermography to extract the temperature and strain gradients across the thickness of a GaN buffer were investigated.

Despite the promise of solid immersion lenses, a spatial filter was developed further as it is more generally applicable. By combining the spatial filter with azimuthal polarisation, and with the use of offset focal volumes to select Raman-scattered light from specific regions of the buffer, it was possible to obtain the temperatures of the top and bottom of a 1.9 µm GaN buffer. The results obtained by this method agreed with previous work for strain gradients.

The temperature gradient across a GaN buffer in an operating HEMT determined experimentally was found to be good agreement with a thermal simulation using a homogeneous GaN thermal conductivity and where the only free parameter is the effective thermal boundary resistance, which itself agrees with previous work. This simulation required the development of a method for spatially averaging the simulation output in a physically realistic manner. Assuming a homogeneous GaN thermal conductivity is in agreement with transmission electron microscopy (TEM) images which tend to show a layer of low-crystal-quality GaN comparable in thickness to the nucleation layer and adjacent to the nucleation layer.[55,187]

A further experiment providing additional confirmation of this crystal quality gradient in a more direct form may be to consider the widths of the GaN spectral lines at the top and bottom of the GaN in an unpowered device.

This new experimental method provides a direct measure of the elevated temperature of the GaN surface in AlGaN/GaN HEMTs and the vertical temperature gradient through the GaN device layer. It furthermore enables thermal simulations to be refined by fitting to improved experimental temperature data in AlGaN/GaN HEMTs.

While new results are available using this method, it does have some drawbacks. The measurements are time-consuming, which would tend to restrict their use to research rather than materials testing, as well as requiring a stable system. It is also only possible to measure areas of the device surface with good optical access, therefore it is not suited to field-plated devices. Devices on Si substrates, which absorb the laser light, would present further challenges, due increased laser-heating of the substrate caused by the higher laser powers used here than in most Raman thermography measurements. A potential solution to the issues caused by the long measurement time may be to use diamond particles as surface thermometers, as described by Simon *et al.*,[188] combined with optimising the measurement depth to give a spectrum weighted towards the bottom

of the buffer. This would take advantage of the non-zero signal even from layers some distance from the focus, and the fact that diamond is a strong Raman scatterer with a peak that does not coincide with any from other materials of interest. This peak is in fact far enough from the peaks due to GaN and SiC that a careful choice of grating and spectrometer configuration would be required. With regard to broader applications of the measurement approach presented here, some constraints will apply to any samples on which it is to be used: (i) they must be reasonably strong Raman scatterers; (ii) they must not have metal contacts (unless so thin as to be transparent) on the surface near which the temperature is to be measured; (iii) ideally, they should consist of an epilayer on a substrate. This last constraint would rule out some of the potential applications, however for samples such as all-SiC or bulk-GaN devices there may still be some benefit at least as regards the ability to obtain a more surface-sensitive temperature than in conventional Raman thermography, as well as a benefit to a well-quantified spot size inside the sample. This could apply to vertical devices, whether GaN-based[189] or SiC,[190] but would be subject to suitable optical access. This could be achieved by masking (at the time of deposition) or etching (on an existing device) a circular region in the centre of the top contact. If the masked region was small compared to the gate, this should have little or no effect on the device operation. The interpretation of results from near any back metal contact would have to be carried out with care, taking into account reflections from the metal.

Chapter 6

Electroluminescence in high electron mobility transistors

The phenomenon of electroluminescence in HEMTs has been used to study various parameters of operating devices, such as the field distribution in the on state,[100] and the presence of impact ionisation in AlGaN/GaN/InGaN DHFETs.[93] Its applicability in the case of off-state stress has also been established in recent years. In this chapter off-state electroluminescence is applied to sites of gate leakage current – "hotspots" – as they form under harsh off-state stress conditions. Photoluminescence spectroscopy is also used, to probe the changes occurring in a device as the result of off-state stress. The application of photoluminescence spectroscopy to devices has generally been as a tool for surface-sensitive thermography, however photoluminescence has long been used to study semiconductor materials, for example to study damage to SiC caused by electron beams.[191] While these changes are detectable optically due to the light they give off, they occur on a scale too small to study with visible light. The degradation also shows poor correlation with some of the candidate mechanisms for the formation of leakage paths: for example more hotspots form along the gate edge than would be expected if they formed at threading dislocations, given the dislocation density. The goal of the combined approaches set out here is to improve the understanding of the processes taking place as these leakage paths form.

Some of the results in this chapter have been published in Reference 104: C. Hodges et al. 'Optical investigation of degradation mechanisms in AlGaN/GaN high electron mobility transistors: generation of non-radiative recombination centers'. *Applied Physics Letters* 100.11 (2012). In addition, some of the methods discussed here (particularly the current and image logging automation) were developed for use in the work described in Reference 192: M. Montes Bajo et al. 'On the link between electroluminescence, gate

current leakage, and surface defects in AlGaN/GaN high electron mobility transistors upon off-state stress'. *Applied Physics Letters* 101.3 (2012), p. 033508.

6.1 Experimental

Before discussing the results in detail it is necessary to establish some experimental techniques, particularly as some of them are newly-developed for the work described here. The basic equipment of spectrometers, microscopes and electrical connections is similar to that already described in Chapter 3; for much of this work a cooled (−10 °C) black-and-white CCD camera was fitted to the microscope.

6.1.1 Hot-electron electroluminescence

Hot-electron electroluminescence, both as images and as spectra, is a useful tool for easily visualising some of the electrical properties of an operating HEMT. As hot-electron electroluminescence emission requires both current and field (see Section 2.5) it is useful for finding the peak field region in an on-state device, and regions of localised current flow – *hotspots* – in off-state devices, simply by photographing the device under a microscope as it operates. A spectrometer may be coupled to the microscope to gain information about the carrier energy distribution, characterised by an electron temperature (Equation 2.51). The experimental configuration for electroluminescence spectroscopy is almost identical to that for Raman thermography, shown in Figure 3.2, with the exception that the Rayleigh filters are removed.

Treating the 2DEG as a classical system[193] allows the observable carrier energy distribution given in Equation 2.51 to be approximated by the simpler Maxwell-Boltzmann distribution, which may be written as

$$f_{\mathrm{MB}} \propto \sqrt{\frac{E}{k_{\mathrm{B}}T}}^{3} \mathrm{e}^{-E/k_{\mathrm{B}}T} \tag{6.1}$$

where E is the energy of the emitted photons, k_{B} is Boltzmann's constant, and T is the electron temperature. The electron temperature may be found by fitting (Figure 6.1) the distribution in Equation 6.1 to the exponential part of the spectrum recorded. It is necessary before fitting to correct the intensity in the recorded spectrum for the

wavelength-dependent system response, as hot-electron EL is detected over a sufficiently large spectral range that the performance of the spectrometer may not be regarded as constant with wavelength (Figure 6.2). In an AlGaN/GaN HEMT hot-electron electroluminescence may commonly be detected over a range of 1.8 eV to 2.4 eV (most of the visible wavelength range). An experiment considering also GaN bandgap recombination electroluminescence would extend this range to 3.5 eV, i.e. into the ultraviolet. Over this range the transmission of the lenses, the reflectivity of the mirrors and diffraction grating, and the sensitivity of the CCD all vary. The system response was calibrated with respect to a known black-body source.

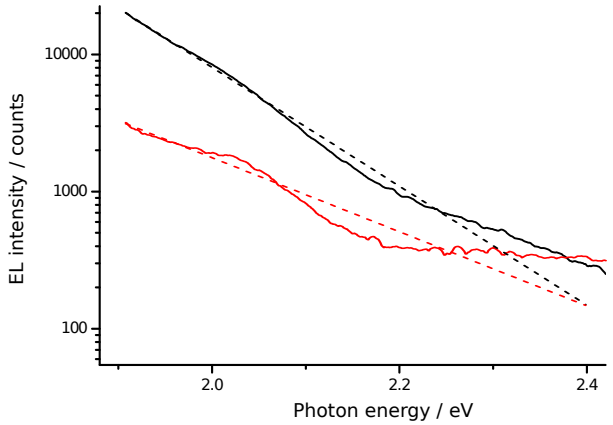


Figure 6.1: Maxwell-Boltzmann distribution (dashed lines) fit to electroluminescence spectra (solid lines) for on-state (black, 1136 K) and off-state (red, 1798 K).

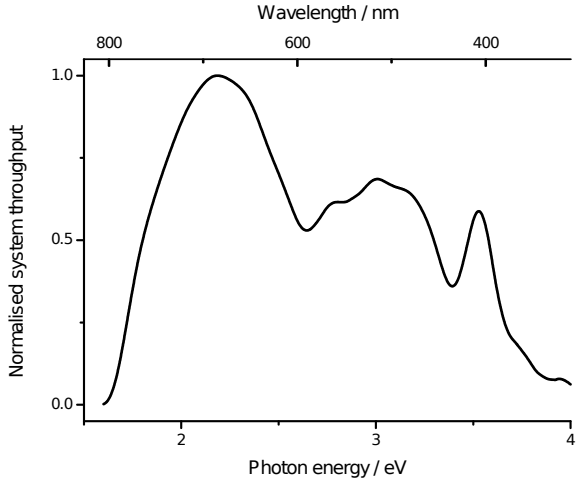


Figure 6.2: The throughput of the 325 nm ultraviolet spectrometer including the detector response, normalised to the peak throughput. The system throughput was found by recording a spectrum of a white light source which had a calibrated output spectrum. The recorded spectrum was then divided by the light source output spectrum. Many long acquisitions were averaged to produce this spectrum with minimal noise.

As the peak electric field is located at the edge of the gate on the side facing the drain,[194] a T-gate or field plate will obscure a significant proportion of the light emitted

in this region from the objective. It is therefore sometimes necessary to measure electroluminescence from the back of the device, through the substrate, for example using a retro-reflector as described in Section 3.1.4.[100] Measuring from the back of the wafer is not without its downsides however. The amount of light reaching the CCD camera is noticeably less when viewed from the back of the wafer, as light is lost in the retro-reflector system. This is due to reflections at the extra interfaces and the long optical path through the glass of the retroreflector. The spherical aberrations caused by focussing through the substrate also reduce the optical spatial resolution, as mentioned in Section 2.3, as well as reducing the measured electroluminescence intensity. These effects may be seen in Figure 6.3 which compares electroluminescence images from the top and bottom of the same device under the same bias conditions and acquisition times. When photographed through the retro-reflector the distribution of light is visibly broader with a lower peak intensity. In addition a ghost image of the electroluminescence is visible at the top of Figure 6.3(a), probably caused by Fresnel reflections within the retro-reflector (Figure 3.5). This device had no field plate or gate overhang, both of which would obscure the light in Figure 6.3(b).

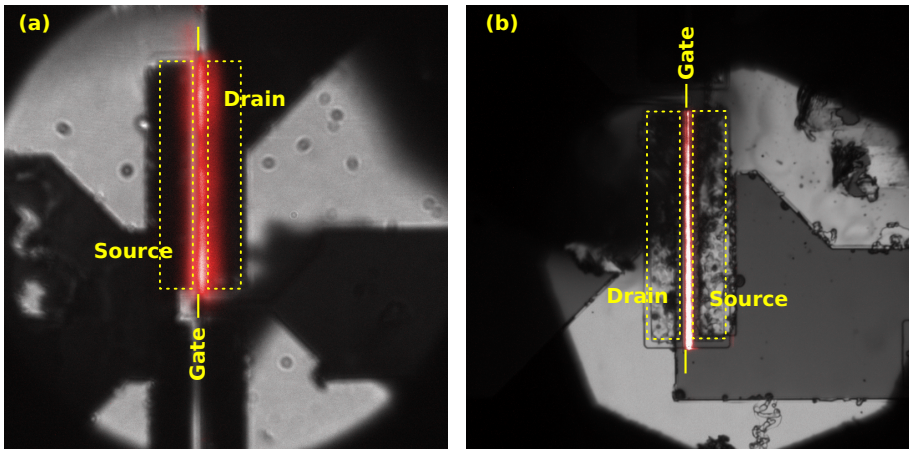


Figure 6.3: False-colour electroluminescence images of a normally-on device produced at UCSB with one finger biased to $V_{DS} = 40$ V, $V_{GS} = 0$ V: (a) viewed through the wafer using a retro-reflector; (b) viewed from the contact face. In both cases the illumination source was above the wafer, equating to transmitted light in (a) and reflected light in (b). The drain current was 55 mA. The loss of imaging quality when viewing through the wafer is clear.

6.1.2 Electroluminescence hotspots

When a HEMT is pinched off, ideally no current would flow at any of the contacts. In practice this may not be true, especially in a device which is degraded as a result of stress. While the off-state currents may be from the source to the drain as in on-state

operation, these are not the dominant leakage currents.[195] Instead current flows from the drain to the gate, despite the Schottky diode at the gate, caused by a degradation of the gate contact. The resulting *gate leakage current* is the topic of much research, as it is implicated in one of the major routes to failure of AlGaN/GaN HEMTs.[196,197] For electrons to flow from the gate to the drain (or indeed to the source) they must pass through the AlGaN barrier to reach the device channel. Some proposed mechanisms for this will be discussed later in this chapter. The leakage current is typically confined to small regions along the drain edge of the gate in the form of hotspots, as may be seen in Figure 6.4, which also demonstrates that these spots appear during stressing. Typically, off-state stress conditions are selected with the intention of causing hotspots to form gradually. To a good approximation it is the sum of V_{GS} and V_{DS} that is important, and for comparison with the same device in the on state it is often convenient to set V_{DS} to a value for which the device is designed, while setting V_{GS} to a more negative value than that at which it would be operated (hard pinch-off).

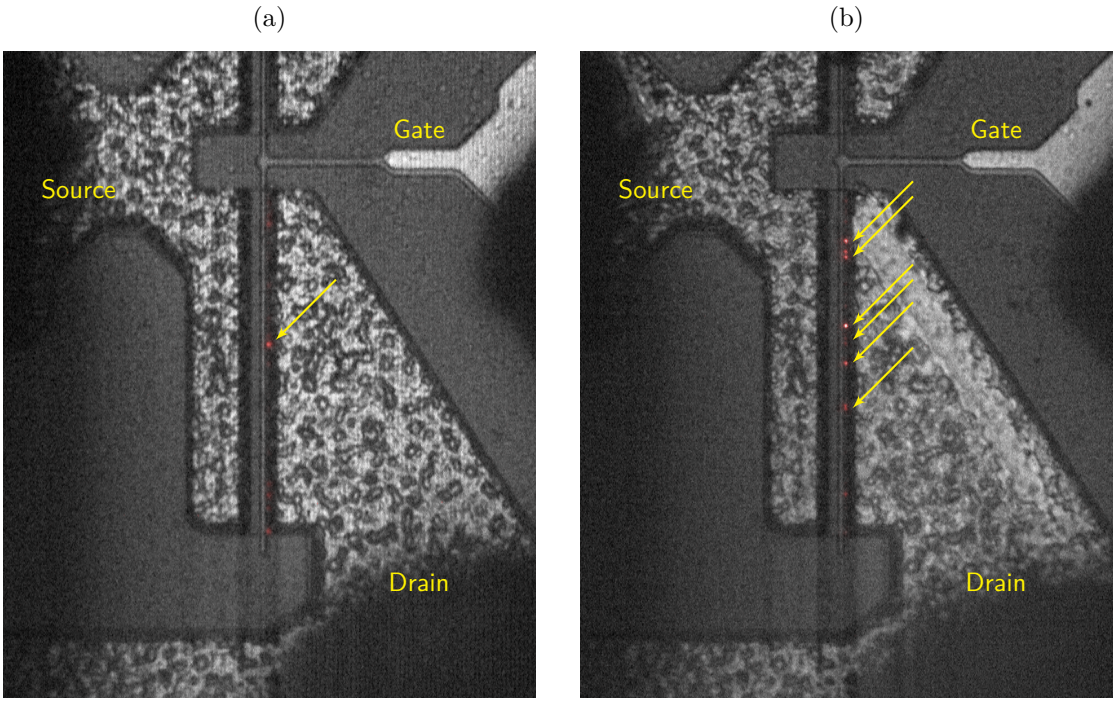


Figure 6.4: Electroluminescence hotspots (coloured red) overlaid on white light images of a device (a) before; and (b) after 2 hours off-state stress at $V_{GS} = -8$ V, $V_{DS} = 12$ V.

To study these hotspots in any detail, for example by photoluminescence,[104] AFM,[192] or TEM[198] requires first that they are located precisely. Optical microscopy is particularly convenient for this as an entire device may be imaged simultaneously and quickly, with high resolution. Automatically logging the development of hotspots is therefore easy,[192]

and this may be correlated to currents in the device as will be shown in Sections 6.2.1 and 6.2.2.

6.1.3 Experimental details

After stressing devices until hotspots formed (20 h at $V_{GS} = -16$ V and $V_{DS} = 25$ V), the locations of the hotspots were compared to other points along the gate edge, which did not emit light, by deep ultraviolet photoluminescence. This was done in order to ascertain whether the changes in the device implied by the formation of hotspots affected the AlGaN, the GaN or both. The devices studied consisted of 900 nm of GaN grown by plasma-assisted molecular beam epitaxy on a SiC substrate, with a 30 nm $Al_{0.30}Ga_{0.70}N$ barrier. The gate width was 75 µm, and the source-drain spacing was 5 µm with a 1.5 µm gate. The 0.75 numerical aperture objective used for the electroluminescence measurements gave an imaging resolution of ~0.7 µm. Off-state electroluminescence spectra were recorded from the entire gate region, including several hotspots, and compared to on-state electroluminescence from the same region.

Devices were also operated inside the Linkam liquid-nitrogen-cooled sample cell. As it was necessary to exclude water vapour and CO_2, which would condense and freeze on the sample surface at low temperatures, the cryostat had to be sealed and therefore it was not possible to make electrical contact to the device under test using probe tips. Instead a sample was prepared by mounting a wafer-piece on a carrier plate, and wire-bonding connections to a printed circuit board, which could then be connected to an electrical feedthrough into the cryostat.

Photoluminescence (PL) spectroscopy on unbiased devices after off-state stress was performed using a frequency-doubled argon-ion laser at 244 nm (5.08 eV) i.e. excitation above the AlGaN bandgap (4.07 eV) for these devices. A 0.5 numerical aperture objective (40×) was used to focus the laser light onto the device and to collect the photoluminescence signal. The photoluminescence spot size was ~5 µm and the laser power incident on the sample was ~0.2 mW.

In order to study the formation of hotspots it was desirable to compare some properties of newly formed and established hotspots. To do this it was necessary to stress a device until the gate leakage current started to increase. For the experiment described here this was done by monitoring the gate leakage current I_G in the device using a computer

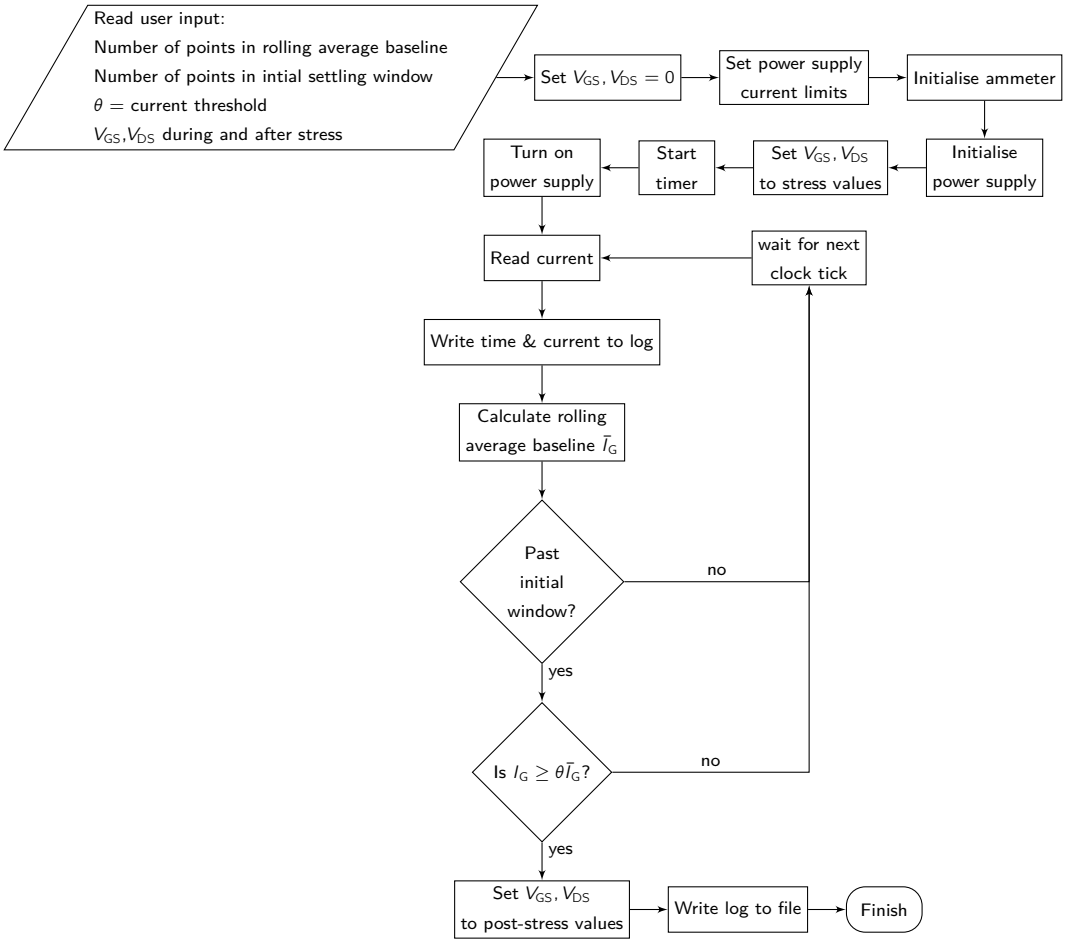


Figure 6.5: The process of automatically stopping device stressing after an increase in the drain current.

which was connected to a digital multimeter and could control the output of a power supply. The computer was also used to record electroluminescence images from a Peltier-cooled CCD camera monitoring the DUT through a microscope. Figure 6.5 summarises the control system as implemented in *National Instruments Labview*, and the hardware configuration is shown in Figure 6.6. The load resistors R_{L1} and R_{L2} were used to ensure rapid discharge of the output capacitors of the power supply when the output was disabled or the output voltage was reduced, i.e. to ensure that the stressing finished when it was expected to.

The process of automatically stopping the stress as shown in Figure 6.5 was complicated by the current profile when the device is first turned on. The gate leakage current flowing in the channel of a device typically shows an initial decrease before forming a stable if slightly noisy baseline. This is due to trapping of electrons under the gate.[199] It is at the first increase from this baseline that the first hotspot forms, as shown in Figure 6.7. It was therefore necessary to consider that a hotspot was likely to have formed when the

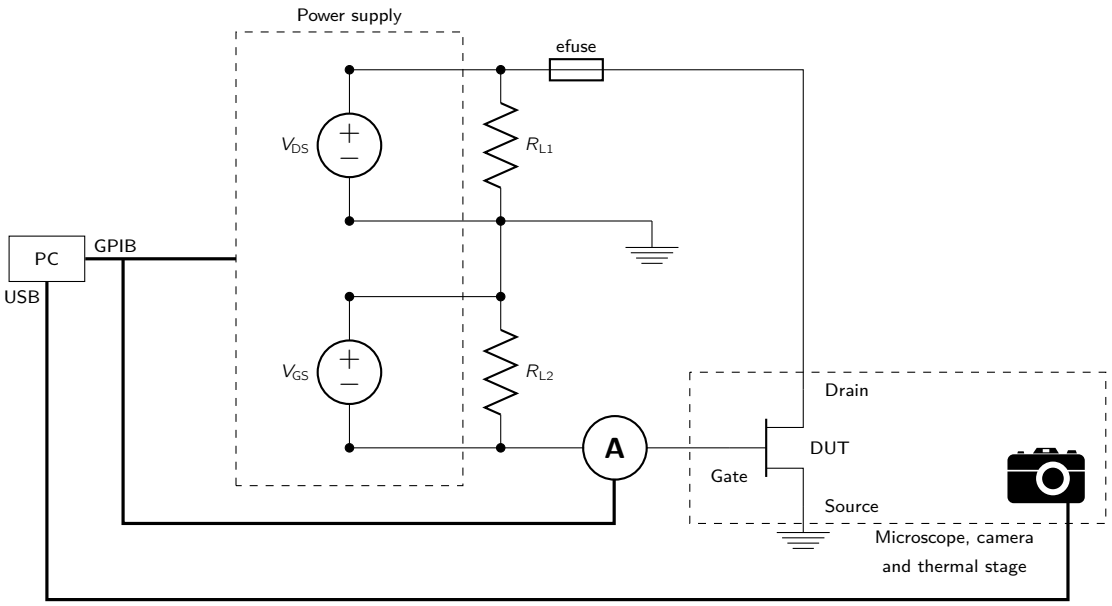


Figure 6.6: Schematic of the automatic stop stress system. While the automatic stop stress system and the imaging software were hosted on the same PC, they were independent of each other.

gate leakage current increased beyond an average of the gate leakage current over the preceding few seconds, $\bar{I}_{\mathrm{G}}$. The threshold for triggering is denoted θ in Figure 6.5. In addition there was an initial settling period within which the system would not trigger, to allow a stable average baseline to be established, and to ignore any transient current at switch-on. The option of setting non-zero post-stress gate and drain voltages was implemented to allow the collection of electroluminescence images after triggering while avoiding further stress: the image acquisition rate was considerably slower than the current measuring rate, meaning that without post-stress image acquisition there would be no guarantee of an image allowing the newly-generated electroluminescence hotspot to be located.

The evolving intensity of an individual hotspot could be taken from the time series of images using the code given in Appendix A.3. Hotspot positions could be determined manually from a post-stress image, and the intensity at those positions integrated automatically in all earlier images. To take account of system drift the code allowed the spot centre position (as found from the brightest pixel within the spot) to move within user-specified limits. Alternatively the intensity over the whole gate region could be integrated for comparison with the total gate leakage current.

Recombination-related electroluminescence, including donor-acceptor pair transitions proposed as an explanation of spectral features observed in off-state electroluminescence,

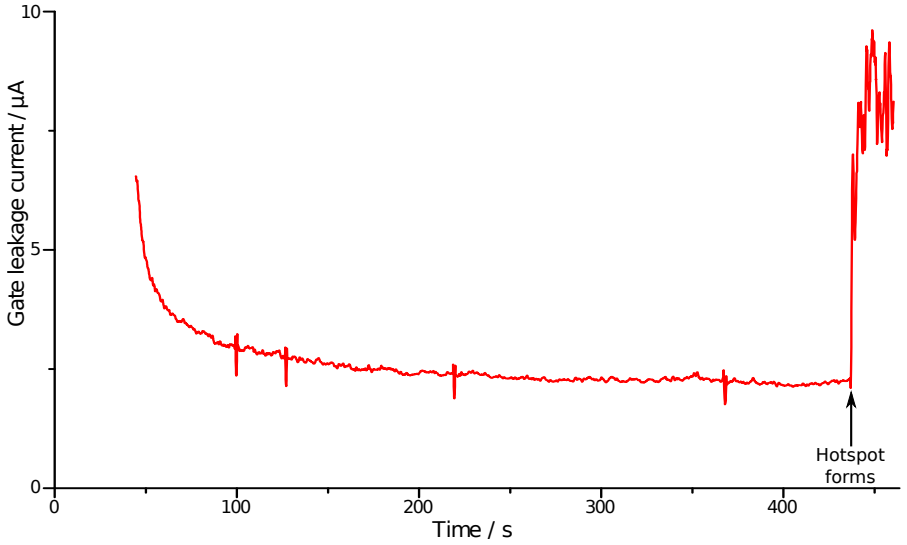


Figure 6.7: The gate leakage current during the early part of the stress shown in Figure 6.8. The first hotspot forms at around 440 s, appearing as a steep rising edge in the current. To detect this rising edge as early as possible it is necessary to take account of the initial decay in the gate leakage current after applying power, which is due to charge trapping under the gate (visible until ~200 s). The spikes in the current are artefacts of the logging system. Figure courtesy of M. Montes Bajo.

requires a source of holes. The main sources of holes in wide bandgap semiconductors are impurity states[200] and impact ionisation.[201] Many impurities have energy levels which are thermally excitable at reasonably low temperatures, even below room temperature, meaning that by operating a device at cryogenic temperatures these energy levels will be "frozen out", and they will be unable to supply holes. Therefore if these levels are responsible for a peak in the electroluminescence spectrum this peak would be expected to reduce at low temperatures.[202] Conversely as the electron mean free path is increased at low temperatures, impact ionisation would be expected to increase when a device is cooled and the supply of holes would increase.[203] By operating a device over a range of temperatures while collecting electroluminescence spectra, it was therefore possible to investigate the source of holes contributing to peaks in the electroluminescence spectrum.

6.2 Results and discussion

6.2.1 Formation of electroluminescence hotspots

The formation and development of hotspots over the course of a stress experiment were investigated by logging the gate leakage current as described in Section 6.1.3 but without automatically stopping the stress, and recording images for analysis using the source

code in Appendix A.3. Figure 6.8 shows the gate leakage current and electroluminescence intensity as a function of stress time. The electroluminescence summed over the hotspots shows a clear correlation to the gate leakage current. In particular rising edges in the gate leakage current correspond to the onset of new hotspots.

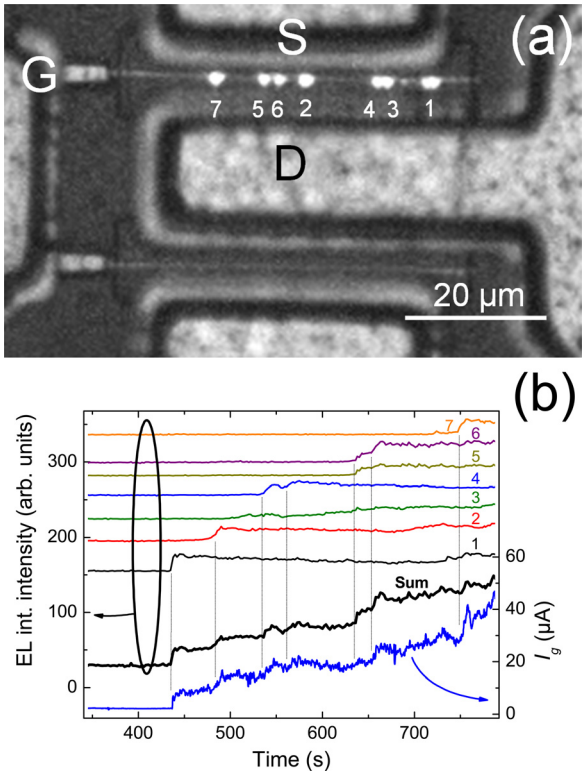


Figure 6.8: The formation of several hotspots in a 50 µm device. One finger was stressed at V_{GS}= −15 V, V_{DS}= 30 V. (a) Electroluminescence (EL) overlaid on a white-light image of the device, the hotspots are numbered in order of appearance; (b) Gate leakage current (right axis) and integrated electroluminescence intensity around each hotspot (left axis, curves offset for clarity). Figure courtesy of M. Montes Bajo,[192] analysis carried out using the code in Appendix A.3.

Figure 6.9 shows one of the devices subsequently studied under deep ultraviolet photoluminescence 30 minutes after the start of stressing and after 80 hours of stressing. At the beginning of the stress period no hotspots were visible. The increase in both number and brightness of hotspots over the stress period is clear, although some hotspots had already formed within the first 30 minutes of stressing.

While the device shown in Figure 6.9 was being stressed, electroluminescence images and spectra were collected, and the gate leakage current was logged. The increase in gate leakage current and the corresponding increase in electroluminescence intensity from hotspots on the drain side of the gate were measured and are shown in the inset to Figure 6.10. In addition to the hot-electron emission (in the range 1.9 eV to 2.2 eV) enhanced emission in the visible spectral region (2 eV to 3 eV) is apparent in the off-state

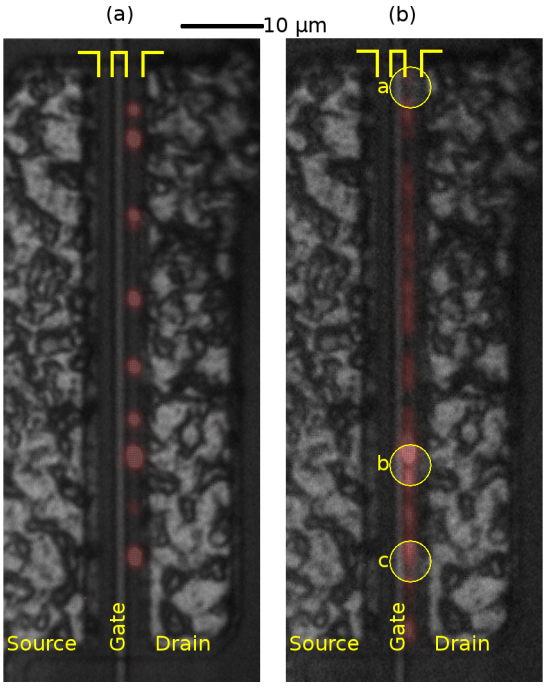


Figure 6.9: Electroluminescence images (coloured red) overlaid on white-light images of an AlGaN/GaN HEMT after stress at V_{DS}= 25 V, V_{GS}= −16 V for (a) 30 minutes; (b) 80 hours. Circles in (b) indicate the location and spatial extent of subsequent 244 nm photoluminescence measurements (Figures 6.12 and 6.13).

spectrum, possibly related to yellow luminescence, i.e. defect-related optical transitions. The ripples visible in the tail of the off-state spectrum are caused by Bragg interference effects due to reflections within the device. The part of the total electroluminescence intensity due to hot electrons increases significantly in the early part of the measurement before saturating, while the electroluminescence in the range 2.8 eV to 3.1 eV shows a slight decrease. This may be an artefact, possibly related to loss of focus over the course of the measurement. The electroluminescence signal in this region was variable from device to device. Similar hotspot spectra were also recorded from metal-organic chemical vapour deposition grown devices; in some cases the non-hot-carrier related feature could even take the form of a distinct peak or peaks at 2.8 eV (Figure 6.11); and in other cases there was no optical emission in this spectral range. In the on-state electroluminescence spectrum only hot-carrier contributions are apparent, and no defect-related features can be seen. The on state spectrum shown was recorded before stress; spectra recorded after stress were similar with the intensity reduced slightly (blue dotted curve in Figure 6.10), as expected from previous work.[204] The tails of the spectra are effectively dependent on the hot-carrier temperature as given by Equation 6.1. The hot electron temperature was determined, from this contribution to the electroluminescence, spectra to be 1534 K in the off state, and 1190 K (before stress) to 1304 K (after stress) in the on state, as may

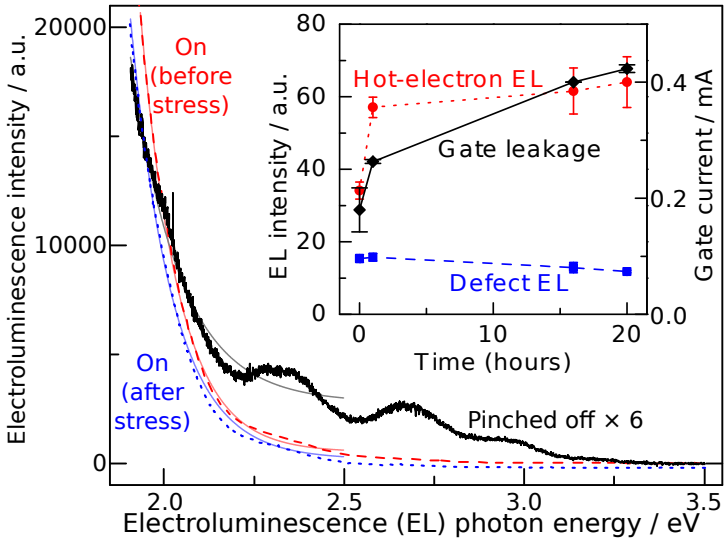


Figure 6.10: Electroluminescence spectrum of a typical AlGaN/GaN HEMT at $V_{DS} = 25$ V in off state ($V_{GS} = -16$ V, solid black line). For comparison EL spectra recorded in the on state are also shown ($V_{DS} = 25$ V, $V_{GS} = 0$ V, dashed red line before stress, dotted blue line after stress). The ripples visible in the spectra are due to interference of light caused by internal reflections within the device (Bragg interference). Light solid lines indicate the Maxwell-Boltzmann distribution fitted to the data. Prior to the measurements the device was stressed for 17 hours at $V_{DS} = 25$ V and V_{GS} up to -16 V. The inset shows hot-electron electroluminescence (red circles, 1.9 eV to 2.2 eV) and defect-related electroluminescence (blue squares, 2.8 eV to 3.1 eV) as a function of stress time, together with leakage current (black diamonds).

be seen in Figure 6.10. This was observed consistently, for example in another device the hot electron temperature determined from the hotspots at $V_{DS} = 25$ V, $V_{GS} = -10$ V was 3100 K. This is higher than the hot electron temperature determined at the same source-drain bias in the on state (1800 K).

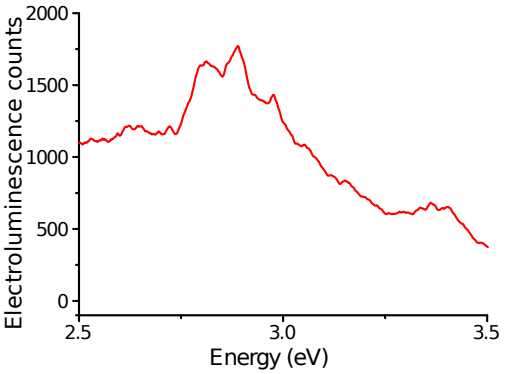


Figure 6.11: Electroluminescence feature observed in a HEMT pinched off at V_{DS}= 60 V, V_{GS}= -40 V. This feature was not observed in on-state, and is probably related to donor-acceptor pair-assisted recombination.[95] The spectral baseline is part of the tail of the hot-electron spectrum.

6.2.2 Photoluminescence studies at hotspot locations

Photoluminescence measurements were performed at sites in the devices where electroluminescence hotspots emerged during off-state device stress, as well as reference points without hotspots, and are illustrated in Figure 6.12 for the locations indicated in Figure 6.9.

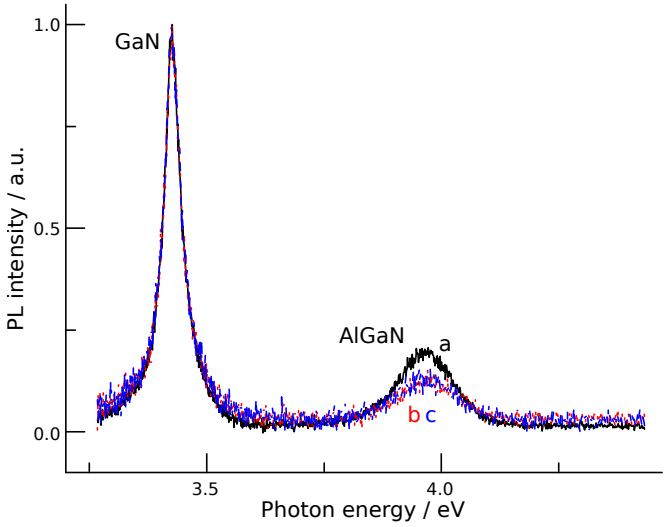


Figure 6.12: Photoluminescence spectra from three regions of the same device: (a) without hotspots; (b) and (c) with hotspots, corresponding to the points labelled in Figure 6.9(b).

Clearly apparent in the photoluminescence spectra are two features, one related to the GaN bandgap (3.4 eV) and one to the AlGaN bandgap (4.0 eV). A significant reduction of the AlGaN bandedge photoluminescence compared to the GaN bandedge photoluminescence is apparent at hotspot locations. This is also illustrated in Figure 6.13 demonstrating an analysis of a selection of hotspot locations, clearly showing that for each spectra recorded in regions with hotspots the AlGaN bandgap photoluminescence is reduced, with a greater reduction corresponding to a higher density of hotspots. With the diameter of the laser spot of ~ 5 µm being larger than the hotspots observed in electroluminescence, this illustrates that at the site of each individual hotspot the AlGaN photoluminescence must be reduced dramatically. The electroluminescence hotspots in Figure 6.9 are not fully resolved, but the smallest spots appear less than 1 µm in diameter.

No significant reduction in the absolute GaN photoluminescence intensity was found, demonstrating that all changes take place in the AlGaN layer, during the device stress observed here. Similarly, purely optical effects caused by changes in the barrier near a hotspot are likely to be minimal, as these would affect the signal from the buffer as well as

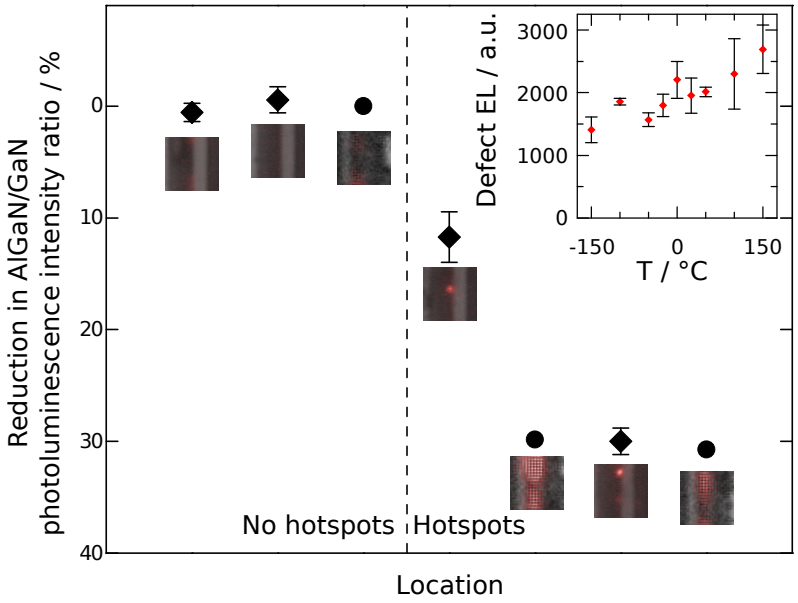


Figure 6.13: AlGaN to GaN photoluminescence intensity ratio (integrated peak area) recorded at selected locations along the gate edge of degraded devices, in regions with and without hotspots. Electroluminescence images of the recorded areas are depicted. The inset shows the electroluminescence (EL) intensity in the 2.8 eV to 3.0 eV spectral range as a function of device temperature.

from the barrier, given that the barrier thickness is $\sim \lambda/10$. As the 244 nm laser is absorbed in the AlGaN barrier and the top of the GaN buffer of the devices, the observed reduction in the AlGaN photoluminescence intensity in hotspot regions therefore indicates the generation of non-radiative recombination centres within the AlGaN barrier layer. In III-V semiconductors point defects have been previously reported as being responsible for non-radiative recombination, such as reducing LED efficiency.[205]

To obtain a better measurement of the form and size of the defect structures underlying hotspots, AFM has been used.[192] AFM measurements of the GaN surface require removal of the contacts and passivation layer by etching, making this a destructive technique. Pits, which appear to form at the top of current leakage paths, were found correlated to the hotspot locations at the gate edge, with a typical lateral size of $\lesssim$200 nm, and depths of a few nm, as illustrated in Figure 6.14. Some pits were elongated towards the drain, while others were not, with variations within a single device as well as between devices and stress conditions. A feature of this size is below the lateral diffraction limit for both white-light imaging and deep ultraviolet photoluminescence. This means that while it may be detected optically by its electroluminescence emission, optical methods such as photoluminescence must be combined with higher resolution methods if the role that these features play in device degradation is to be understood. As the pits only account for a small proportion of the thickness of the barrier (of the order of 10 %, which is two

orders of magnitude less than the laser wavelength), the reduction in photoluminescence signal due to the laser encountering a reduced volume of material may also be expected to be small, and the generation of defects within the AlGaN barrier must be the origin of the dramatically reduced photoluminescence signal.

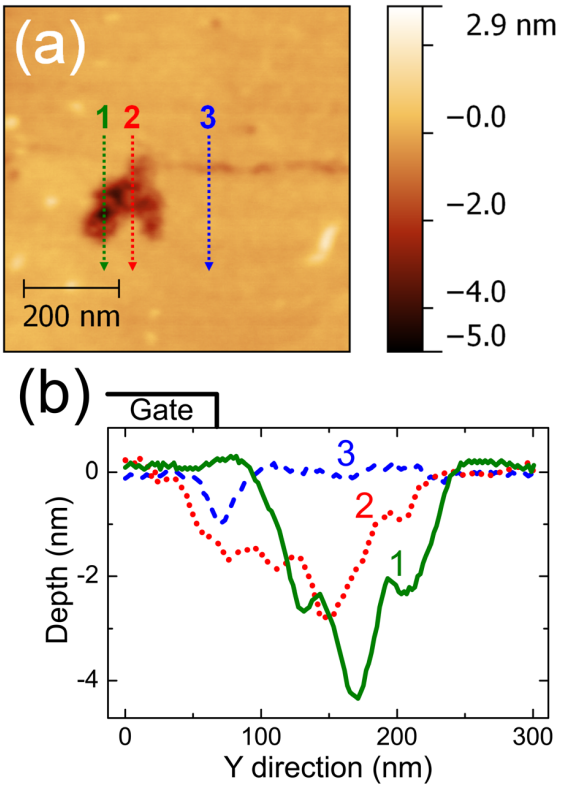


Figure 6.14: (a) AFM image of a surface pit formed during off-state stress. (b) AFM Depth profiles along arrows marked in (a). Figure courtesy of M. Montes Bajo.

A typical lateral pit size is compared to the photoluminescence laser spot in Figure 6.15, which indicates that the pit makes up a small fraction of the area measured by photoluminescence. The fraction of the photoluminescence laser power incident on the pit was larger than the fraction of the beam area occupied by the pit, due the the distribution of power in the laser beam, which was centred as closely as possible on the hotspot position. This power fraction was determined by numerical integration assuming a Gaussian laser spot. Of the photoluminescence laser power which is incident on the semiconductor surface and therefore may give rise to a photoluminescence signal (i.e. which does not encounter the metal contacts) less than 1.6 % encounters one typical pit (neglecting for the moment carrier diffusion lengths and treating the pits as circular). It is reasonable to assume that spots which appear elongated parallel to the gate in Figure 6.9(b) contain multiple pits, i.e. are actually multiple hotspots with multiple underlying pits. In the limiting case of a continuous damaged region extending 250 nm from the gate

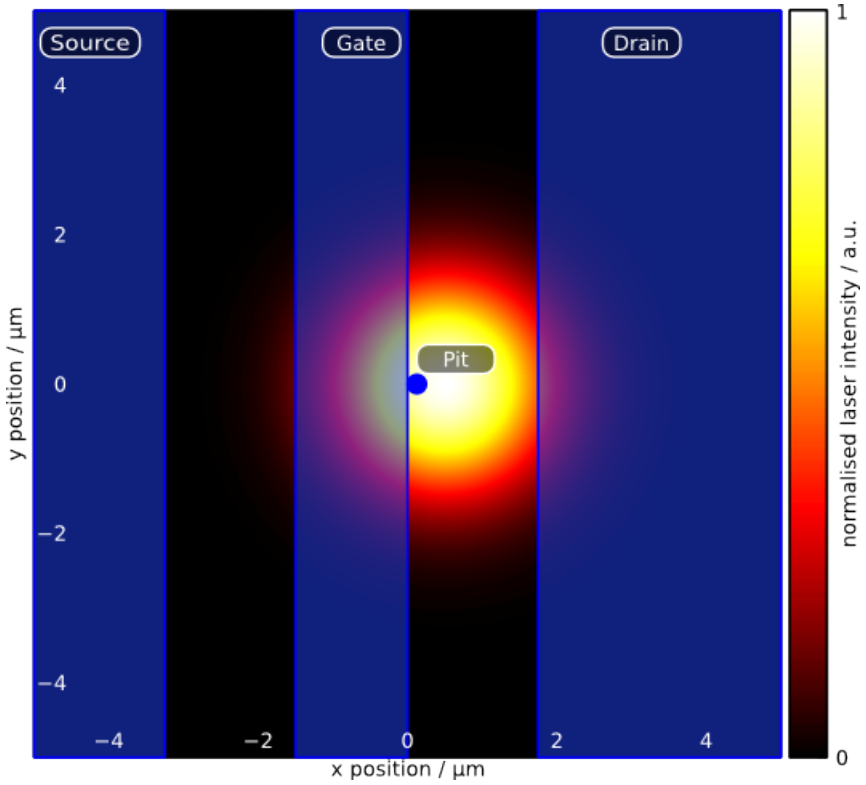


Figure 6.15: The relative sizes of the Gaussian 244 nm laser spot and a typical pit found at a hotspot location[192] shown as a blue dot. The blue shaded regions indicate the extent of the contacts.

edge towards the drain (a much higher damage density than that reported)[192] 28 % of the laser power incident on the semiconductor surface encounters the damaged region. The reduction in photoluminescence peak intensity is larger than this – around 50 % – so even if the non-radiative recombination centres created at the hotspots could be characterised as completely efficient and covering the whole area of the pit, they would not be able to account for the reduction in the AlGaN barrier photoluminescence signal.

It is likely for the influence of the defects to extend further laterally, in particular by drift or diffusion of carriers, i.e. for carriers to move towards the defects. Hole transport in semiconductors such as AlGaN may also be enhanced by "hopping" of holes between states, which could potentially enhance the capture of holes by non-radiative defects. Considering the opposite limit, of a single cluster of defects at the site of the electroluminescence spot causing the non-radiative recombination of sufficient electron-hole pairs to account for the observed decrease in photoluminescence from the barrier, a defect would have to have a region of influence with a radius of ~1.2 µm and 100 % capture efficiency. In the absence of a lateral field (the built-in fields shown in Figure 2.5(b) are vertical) during the unbiased photoluminescence measurements, field-driven drift

can be neglected, but diffusion must be considered. Minority carrier diffusion lengths from the literature must be used with care – very little work has been reported on diffusion lengths in AlGaN – however it is still possible to estimate the effective size of the diffusion-limited region. Reported values for the minority carrier diffusion lengths in III-nitrides vary from ~50 nm[206,207] to greater than 1 µm.[208,209] In particular for the GaN in an InGaN/GaN heterostructure 1.7 µm has been reported.[210] While the upper limit of the reported values is sufficient to allow a single damaged region within the focus to explain the reduction in photoluminescence, it is not possible to conclude that this is the whole explanation for several reasons. These larger figures for the hole diffusion length are for GaN rather than AlGaN, and either for nanowires,[209] or n-doped GaN, making a direct comparison to undoped AlGaN unreasonable.[208] The lower literature values are generally for InGaN, and are thus likely to be more comparable to AlGaN given that alloy disorder reduces the mobility in ternary semiconductors. Taking a typical lower figure from the literature of an additional 250 nm diffusion beyond the measured pit size allows the hotspot density that would have to exist at the laser focus (assuming even spacing along a line parallel to the gate edge) to be calculated. The maximum separation between regions with a high density of non-radiative recombination centres would have to be no more than ~0.5 µm, comparable to the saturated hotspot density found by Montes Bajo *et al.*[211] after many hours of harsh stress conditions. As demonstrated by the presence of areas free from hotspots, the hotspot density in this experiment should not be regarded as having saturated. It is therefore probable that carrier diffusion to a cluster of defects at the location of a pit could account for a large part of the reduction in photoluminescence observed. It is not possible however to be certain that this is the whole explanation, as it relies on both a high density of pits and efficient capture of generated electron-hole pairs.

6.2.3 Hotspot formation processes

Based on the experimental results described above, a model for electroluminescence hotspot generation can be developed. Random trap generation during stress, forming percolation paths through the barrier, has been proposed to account for electroluminescence hotspots[212] in a similar manner to the degradation of oxide layers in Si devices.[213] As the defects underlying these traps are proposed to be formed by the high field present near the gate edge (either directly or through the inverse piezoelectric effect) they are

likely to be distributed randomly along the gate. It is only when a defect path comprising multiple defects connects the gate to the 2DEG that a leakage current can start to flow, and electroluminescence is subsequently observed as the electrons are accelerated towards the drain. The effect of this is that defects, which can form non-radiative recombination centres, will exist at more locations than just at hotspots, with varying density. As a high (generated) defect density is a step towards the formation of the actual hotspots, it might be expected that a high density of non-radiative recombination centres, which are not directly observable, correlates to a high density of hotspots. It may be possible as a future experiment to study this further by mapping the AlGaN:GaN photoluminescence intensity ratio on a fresh device, and comparing this to a map on the same device after stress. The acquisition time (of 30 s per spectrum), combined with the need for up to 100 spectra to one map of a single device depending on the gate width for the deep ultraviolet photoluminescence would make this a rather lengthy experiment requiring a stable system.

The emerging hotspots, which are related to leakage currents generated during device stress, give rise to hot-electron electroluminescence when electrons are injected through the AlGaN barrier into the channel, where they are subsequently accelerated by the electric field along the device channel and flow to the drain. This emission is apparent in the hot-carrier emission tail observed in the electroluminescence spectrum (Figure 6.10). Its intensity is correlated to the gate leakage current through each particular hotspot (Figure 6.8), i.e. is a measure of the local leakage current, with the total leakage current being the integral over all hotspots, as the leakage current prior to the formation of hotspots is typically negligible, as is the electroluminescence. The gate current increases during device stress at constant bias as more leakage paths are formed. Under the stress conditions used for this work the drain current increased with the gate current; no increase in leakage from source to drain was observed, unlike in some step-stress measurements.[214] Photoluminescence spectra showed the generation of defect states, namely non-radiative recombination centres, only in the AlGaN barrier at the location of hotspots where leakage current occurs. This strongly suggests that these non-radiative recombination centres provide a percolative leakage pathway from the gate into the channel, as they preferentially occur in the vicinity the hotspot sites.

This conclusion is consistent with the previously reported time-dependent generation of the leakage currents by Marcon *et al.*[215] Mechanisms that have been proposed for

the generation of new defect states during device stress include for example oxygen or carbon related centres,[216] as well as dislocations,[45] and defects generated by piezoelectric stress in the high-field region.[197,217] These defects also affect the on-state drain current by charge trapping, reducing the output power.[216,218]

The higher temperature of the hot-electrons in pinched-off devices is expected as the field is higher in the off state under the same source-drain voltage but different gate bias between off- and on- states. In addition the electron MFP is longer in the absence of self-heating. The defect-related features in the electroluminescence spectrum were found to be more or less constant during device stress. The observed defect feature in the EL spectrum therefore seems not to be related to newly-generated defects, as otherwise an increasing intensity with increasing leakage current would be expected to occur. Instead it is likely to be due to pre-existing defects in the device layers, unaffected by the degradation processes. Features such as yellow luminescence (which has been attributed to Ga vacancies)[219] or shallow-donor deep-acceptor pair recombination in the GaN buffer[95] are known to occur in this spectral range and are likely contributors to this defect emission observed. While only observed in the off state it is not possible to conclusively rule out the presence of such features in the on state as they may be too weak to be observed on the tail of the hot-electron spectrum which is much stronger in the on state.

The existence of defect-related features in the electroluminescence spectrum requires holes to be present in the devices to recombine with the electrons injected from the gate into the buffer. Taking into account the optical throughput of the spectrometer, it is possible to estimate the hole current from the magnitude of the defect-related electroluminescence, To do this it is necessary to consider: (i) the fraction of the gate area that can be observed by the objective; (ii) the solid angle over which the objective can capture light (including the effects of refraction and Fresnel reflection at the sample surface); (iii) the reflectivity of mirrors within the system, and the transmissivity of lenses; (iv) the efficiency of the diffraction grating; and (v) the quantum efficiency of the CCD used to collect the electroluminescence spectra. Taking all of these into account the integrated defect-related electroluminescence signal is equivalent to a hole current in the range of 10^{-17} A to 10^{-16} A. If these holes were supplied by traps, freeze-out would be expected to reduce the defect-related EL emission at low temperatures, while impact ionisation would lead to an increased hole supply at low temperatures due to

an increased electron mean free path.[202] Electroluminescence spectra were therefore acquired at cryogenic temperatures as described in Section 3.4. The inset to Figure 6.13 shows the behaviour of this peak as a function of temperature. The decrease in EL intensity with decreasing temperature suggests that hole traps are the predominant source of holes for the devices considered here, although the detailed physical nature of these hole traps is not clear, considering their low density as demonstrated by the low hole current.

6.3 Conclusions

In order to study degradation processes related to gate leakage currents, HEMTs were subjected to off-state stressing which was demonstrated to generate regions of localised electroluminescence emission. A method was developed to correlate the formation of these hotspots to the increasing gate leakage current during device stressing. The resulting hotspot electroluminescence was studied spectroscopically as well as by imaging. In addition to a higher electron temperature in pinched-off devices, features related to pre-existing defects were observed in off-state spectra from some devices but not in on-state spectra.

Subsequent deep ultraviolet photoluminescence measurements on stressed devices showed that these hotspots corresponded to regions in which the barrier photoluminescence signal was attenuated with respect to areas that did not emit measurable amounts of electroluminescence in the off-state, indicating a significant increase in the rate of non-radiative recombination, over an area greater than the dimensions of typical examples of surface damage at hotspot locations. The formation of non-radiative recombination centres is therefore linked to the generation of defect pathways which allow conduction from the gate to the 2DEG through the AlGaN barrier. This indicates that the formation of non-radiative recombination centres is linked to the generation of leakage paths, however it is possible for non-radiative recombination centres to be generated without forming leakage paths.

The work published in References 104 and 192 and discussed in this chapter indicates that damage to the surface of the barrier is of critical importance to off-state degradation, and therefore protecting the surface is important. Passivation layers must already be optimised to control the surface states, but also play a role in preventing surface

damage, for example due to oxygen. The work described in Reference 192 has been continued (Reference 211); both trap-based leakage and migration of electrochemically-reactive species are implicated in the formation of leakage paths, with the result that control of defect- and impurity-related trap states is required, along with further understanding and control of electrochemical processes on the device surface. As this degradation is field-driven, it may to some extent be managed by the use of field plates; as yet it is uncertain whether field plates can provide a complete solution to leakage path generation.

Chapter 7

Conclusions

Many important developments in the history of semiconductors have been driven by the desire for improved communications; more recently, power conversion and more efficient lighting applications have also driven investment in new technologies. New materials, as well as new device structures, have played a major part in this development – the III-nitride devices studied in this thesis may be thought of as representing both. The benefits of AlGaN/GaN electronic devices are demonstrable, but so are the hurdles that must be overcome before their promise can be realised. It is only in a limited range of applications that AlGaN/GaN transistor technology currently has a clear benefit over competing technologies when all factors including total system cost and ease of integration, as well as high performance and reliability, are taken into account. As well as development of amplifiers working at hundreds of GHz, devices are currently being developed with the intention of delivering normally-off high-voltage switching on Si substrate, leading to some of the devices studied here.

In this thesis, optical and spectroscopic methods were used to study some of the processes involved in device degradation. The thermal behaviour of devices, which is important for on-state degradation processes, was studied by means of Raman thermography and photoluminescence thermography. Thermal simulations were carried out to match the experiments, and analysed using an averaging technique developed to closely match the distribution of laser light in the sample. Electroluminescence (imaging and spectroscopy) and photoluminescence spectroscopy were used to study off-state degradation, a process which depends on the electric field, and leads to localised current paths.

Both the need for normally-off operation and the goal of commercial devices on Si substrates led to the device studied in Chapter 4. This device had a layer structure forming a double heterostructure field effect transistor (DHFET), improving the electron confinement; the AlGaN layers between the substrate and the channel were also designed

to relieve strain in the 150 nm GaN channel, improving its crystal quality and therefore the device performance. The thermal behaviour of this structure was studied in this work. The differing AlGaN compositions proved to be spectroscopically addressable, allowing the temperature as a function of depth to be measured in a single point measurement, with a depth resolution equal to the layer thicknesses. Alternatively a two-dimensional temperature profile could be produced from a one-dimensional spectral map. Analysing these measurements required the development of a method for finding the positions of the Raman peaks within spectra consisting of many peaks, small strong and some weak, on a variable background. The extra thermal information obtained by the use of the layer structure as a probe of the temperature variation with depth was combined with thermal simulation to find the thermal conductivity of the GaN channel layer. This proved to be considerably lower than that of thicker GaN epilayers (60 $\mathrm{W\,m^{-1}\,K^{-1}}$ instead of 160 $\mathrm{W\,m^{-1}\,K^{-1}}$). This result is attributed to the effect of confining the phonon mean free path (MFP) to within the 150 nm GaN channel layer thickness. The AlGaN intermediate layers already have a much lower thermal conductivity than the GaN, making the GaN act as a heat spreader plate. Such a thin GaN layer is therefore a poor heat spreader, an effect which will only become worse as devices with thinner channels are developed. This has implications for the power density that can be supported by a device using such a construction, as the heat extraction to the substrate (and onwards to the package and heatsink) is restricted by the reduced thermal conductivity closest to the source of heat.

While the novel structure of the device studied in Chapter 4 fortuitously allowed spectroscopic addressing of different depths, this did not help with measuring the vertical temperature profile in a more conventional device, with the typical 1 µm to 2 µm buffer thickness. The limits on the depth resolution that can be obtained by Raman thermography are optical. A number of constraints contribute to a typical Raman thermography experiment on a typical AlGaN/GaN high electron mobility transistor (HEMT) giving only a single GaN temperature at a given point on the device – an average through the depth of the buffer. The two most important optical constraints are the diffraction limit and spherical aberration due to refraction at the device surface. This imposes a limit considering the significant vertical temperature gradient through the GaN buffer. This average is weighted according to the laser intensity distribution, which depends mainly on the wavelength of the light and the objective numerical aperture, given fixed material parameters. Some of these constraints are fundamental parameters of the materials and

device structures present in HEMTs, such as refractive indices. Others are related to the more practical aspects of a realistic experiment, for example the need for sufficient working distance between the objective and the device to allow electrical contact to be made. These restrictions were circumvented by optimising the axial resolution of the spectrometer used for Raman thermography, combined with offsetting the focus of the Raman laser to measure spectra weighed towards the top and bottom of the 1.9 µm GaN buffer of a conventional HEMT. The validity of this method was tested by measuring the difference in strain across the thickness of a GaN buffer in a transfer length method test structure. To compare these results to simulations, a method was developed for spatially-averaging the simulation output in a physically realistic manner. Applying known material parameters to this model allowed the simulation to be fitted to the experimental data, given a single thermal boundary resistance. This thermal boundary resistance must take into account not just the thermal resistance of the nucleation layer and phonon scattering at the buffer-nucleation layer and nucleation layer-substrate interfaces, but also the increased thermal resistance of the portion of the buffer closest to the nucleation layer, where the crystal quality tends to be poor. This thermal boundary resistance was found to be a good match to previously reported values. Fitting a single thermal boundary resistance and a single GaN thermal conductivity to the experimental data indicates that this layer of poor thermal conductivity GaN is thin, which matches previously published transmission electron microscopy (TEM) results. The extra temperature data provides a more accurate and more direct measurement of the peak channel temperature, and provides an additional constraint on thermal models.

Off-state degradation is known to lead to hotspots, where current leaks from the gate to the channel. The formation of these hotspots was correlated to the increase in current flowing in the device, by combining image capture and automated analysis with electrical logging. A system was developed to terminate the stressing process after the formation of a single hotspot, as indicated by an increase in the gate leakage current. Post-stress photoluminescence measurements showed a reduced signal from the AlGaN barrier at hotspot locations, demonstrating the formation of non-radiative recombination centres in the barrier. The magnitude of the reduction in the photoluminescence signal compared to the size of damaged regions of the barrier measured by atomic force microscopy (AFM) indicates that non-radiative recombination centres are likely to be generated over a larger area than the surface damage would indicate, despite carrier diffusion allowing photo-excited electron-hole pairs to migrate to the non-radiative centres.

It is clear from this work that the material properties of III-nitride epilayers have a direct effect on the lifetimes of AlGaN/GaN HEMTs. The thermal conductivity of the GaN buffer and the damage response of the AlGaN barrier to high fields are important to significant but different degradation mechanisms. As this thesis has demonstrated, optical methods may be used to investigate both the thermal and the field-driven degradation processes taking place in the epilayers of a device. In older devices, the thermal conductivity of the buffer was restricted by the quality of the GaN growth; in devices with thin buffers ($\lesssim$ 200 nm) it is the fundamental behaviour of phonons that imposes the limit. Epilayer qualities are continually improving, especially for growth on Si as strain relief layer structures are optimised. The parameters studied in this thesis, which are perhaps more fundamental to the materials involved than some other parameters such as dislocation density (which is affected by the growth conditions) will increasingly set the limits on device performance as the growth improves. Multiple field plates may be used to limit the electric field experienced by the buffer, especially when devices are designed for high voltages. While this approach improves the reliability of devices, it adds complexity and therefore cost to device design. For those studying degradation mechanisms it complicates matters, meaning that as device structures are optimised, optical techniques will increasingly need to be applied from the back of the wafer, at least when transparent substrates are used. When devices are grown on opaque Si, this is not an option – high voltage devices on silicon substrates will be challenging to study optically, and they are under active development as they are attractive to device manufacturers. Rather than a direct measurement of the peak channel temperature of such devices, it may be necessary to use a thermal simulation extrapolating temperatures from regions of the devices where optical access is possible. The methods developed in this work to obtain the device temperature as a function of depth will be helpful in such cases, as the extra measured inputs to the model will help constrain it. This is true whether the temperature is that of the lower regions of the GaN buffer, or of strain relief layers between the channel and the substrate. Electroluminescence hotspots in the off state may also be measured through SiC wafers, but not Si wafers. In this case there is no possibility of extrapolating from other device regions, so an understanding of the mechanisms underlying off-state degradation are important. This is an active area of research, employing both the non-destructive methods which this thesis concentrated on and destructive techniques such as AFM.

Despite the challenges of device degradation, heteroepitaxy and normally-off operation

which have been a recurrent theme of this thesis, the ongoing commercialisation of AlGaN/GaN HEMTs serves as evidence of the positive outlook for this technology. Many of the difficulties, while not completely overcome, have been reduced to manageable levels, though a deeper understanding of the parameters which limit device performance and lifetime is required in order for improvements to continue and for devices to reach the levels of performance that the III-nitride materials used in them should make available. The optical methods for studying processes in devices that this thesis has examined are themselves not without challenges, however they have a proven track record of contributing insights into device behaviour, and this looks set to continue.

7.1 Suggestions for further work

Further development of the methods introduced in this work is possible, and may be necessary for the next generation of devices. The methods developed in Chapter 4 will form an important tool for those who study heat flow in the GaN-on-Si power devices currently being developed – and this is an important part of GaN device development at the moment. It is likely that the peak-picking approach will have to be developed further. This development might involve local curve fitting, which may be able to reduce the need for manual input. Some means of removing a device-related background would clearly help with this; a promising approach might be to curve fit the peaks in a very low noise spectrum and subtract the fitted peaks from the original data before smoothing.

The method described in Chapter 5 for improved axial resolution, while in many ways the culmination of this work, is not an easy technique to apply, and can therefore only have niche applications. It may however be possible to combine it with diamond micro-thermometers to acquire a signal in much less time, and it could prove useful on vertical devices subject to suitably-modified contacts. As a research tool, taking the work carried out here and combining it with oil-immersion lenses would improve the axial resolution significantly as the refraction at the surface would be reduced. For thicker GaN buffers this could allow improved temperature mapping through the buffer.

The work described in Chapter 6 has already been developed further – some of it was published in 2012,[104,192] allowing it to influence more recent studies. A faster automated stop-stress system could allow the very newest hotspots to be studied; the measurement of the current, the detection of the rising edge in software (or directly in hardware)

and the turn-off of the power supply would all have to be faster for any significant improvement, with a total speed increase of at least two orders of magnitude. Mapping the photoluminescence spectrum of a fresh device, then stressing and looking for correlations between the pre-stress AlGaN/GaN photoluminescence peak intensity ratio and the locations of hotspots would be of interest, particularly in the light of recent intriguing suggestions that pre-stress off-state electroluminescence imaging may be able to predict at least some hotspot locations.[220]

Publications and Presentations

Journal articles

C. Hodges, N. Killat, S. W. Kaun, M. H. Wong, F. Gao, T. Palacios, U. K. Mishra, J. S. Speck, D. Wolverson and M. Kuball. 'Optical investigation of degradation mechanisms in AlGaN/GaN high electron mobility transistors: generation of non-radiative recombination centers'. *Applied Physics Letters* 100.11 (2012). DOI: 10.1063/1.3693427.

> For this work I planned, with input from co-authors (M. K., D. W.), and carried out the experiments. I analysed the data with advice and support from N. K. and M. K., and produced the first draft including all figures before input from co-authors allowed me to produce the final version.

M. Montes Bajo, C. Hodges, M. Uren and M. Kuball. 'On the link between electroluminescence, gate current leakage, and surface defects in AlGaN/GaN high electron mobility transistors upon off-state stress'. *Applied Physics Letters* 101.3 (2012), p. 033508. DOI: 10.1063/1.4737904.

> My input to this paper was mainly in the design and programming of both the automated data-collection (Section 6.1.3) and the hotspot image analysis (Appendix A.3).

C. Hodges, J. Anaya Calvo, S. Stoffels, D. Marcon and M. Kuball. 'AlGaN/GaN field effect transistors for power electronics – Effect of finite GaN layer thickness on thermal characteristics'. *Applied Physics Letters* 103.20 (2013). DOI: 10.1063/1.4831688.

> In this case I again planned the experiments with input from M. K., S. S. and D. M., and carried them out. The analysis of experimental results was mine (including Figures 2 and 3), as was the first draft with the exception of the theoretical treatment of the thermal conductivity (J. A. C.), and the final version with the support of the co-authors.

A. Schneider, D. Wolverson, K. Sebald, C. Hodges, M. Kuball and T. Voss. 'Analysis of strained surface layers of ZnO single crystals after irradiation with intense femtosecond laser pulses'. *Applied Physics Letters* 102.21 (2013), p. 211904. DOI: 10.1063/1.4807923.

> Here my input was mainly in the form of assistance with the Raman measurements, including planning.

C. Hodges, J. Pomeroy and M. Kuball. 'Probing temperature gradients within the GaN buffer layer of AlGaN/GaN high electron mobility transistors with Raman thermography'. *Journal of Applied Physics* 115.6 (2014). DOI: 10.1063/1.4865296.

For this paper I was responsible for both the theory and the experimental work, with which I had support from both co-authors. The analysis and first draft were my work, but the input of both co-authors on the design of Figures 3 and 4 was significant. After input from both co-authors I was responsible for the final version.

J. Möreke, C. Hodges, L. L. Mears, M. J. Uren, R. M. Richardson and M. Kuball. 'Liquid crystal electrography: electric field mapping and detection of peak electric field strength in AlGaN/GaN high electron mobility transistors'. *Microelectronics Reliability* 54.5 (2014), pp. 921–925. ISSN: 0026-2714. DOI: 10.1016/j.microrel.2014.01.006.

I was responsible for much of the optical design in this work, as well as having an input into the planning of the experiment and the image analysis.

Conference presentation

C. Hodges, N. Killat and M. Kuball. *AlGaN/GaN HEMTs Device Degradation: Hot Carrier and Blue Defect-Related Electroluminescence*. Glasgow: 9th International Conference on Nitride Semiconductors, 2011.

As well as delivering the presentation, I was responsible for the underlying data collection and analysis (with input from co-authors on both the design and the analysis of the experiments). All figures (apart from one from the literature) were mine.

Poster

C. Hodges, J. Anaya Calvo and M. Kuball. *Improving depth resolution of Raman thermography on GaN HEMTs: Temperature in GaN channel and temperature gradients in the GaN buffer layer*. Bristol: UK Nitrides Consortium, Jan. 2014.

This poster was based on work from two of my papers, so the experimental results shown were mine, as was the poster itself, though some of the figures had contributions from co-authors of the papers.

Appendix A

Analysis source code

A.1 Peak picking code

The Python code set out in this section was written for finding the peaks from the GaN and $Al_xGa_{1-x}N$ epilayers of the devices studied in Chapter 4. It was also used to find the positions of both Raman and photoluminescence peaks in some of the other results in this thesis.

```
# Chris Hodges, CDTR group 2012
#use WiRE's batch file converter tool first to convert to .txt.
#note - x and y have 2 meanings in the sense of mapping, xpos, ypos etc. are spatial coordinates.
import ConfigParser, re, time
import sys, os, glob
from math import *
from numpy import *
from sgfilter import *
from pylab import *
from scipy import *
from scipy.interpolate import interp1d

def vprint(ToPrint):
  if verbose:
    print ToPrint
  messagefile.write (str(ToPrint)+"\n")

def ShowHelp():
  print('''PeakPick.py
==========
Written by Chris Hodges, CDTR group 2012
Loads a txt file containing 1 or more spectra & finds peaks in region of interest.  ROI, smoothing parameters
 etc are set in peakpick.cfg.

Brief (unless "verbose") output is to STDOUT. Specifying "file", "graph" or "summary" on the command line, or
 setting Always_file_output, Always_summary or Always_graph in PeakPick.cfg  gives useful output to files.

Command line options:
====================
help, --h, ?:    Show this help
setup:           Create the config file with default values
verbose, --v:    Show lots of output
graph, --g:      Produce graphs (for debug)
file, --f:       Output to files at various stages. <filename-PEAKS.ext> contains the filtered list of peaks (
 if filtering); <filename-all-peaks.ext> contains a list of all peaks.
summary, --s:    Output summary data.
background, --b: Subtract background (specify file in peakpick.cfg).  Must be in the form x<tab>y or x,y (no
 map coords) on the same x grid as data, smoothed if required. Interpolation uses settings as for spectra.

Anything else is interpreted as a filename & should be last argument.
Note, it's possible to force some of these options in peakpick.cfg
Parameters related to smoothing, differentiation and picking are set in peakpick.cfg.

The process is as follows (config file parameters in square brackets for each step, with suggestions):
  * The file is tested to see if it is a single spectrum or a map.
  * A spectrum is loaded.
```

```
  * Spectrum is smoothed. [sg_window 9, sg_order 2]
  * Spectrum is interpolated to finer steps. [interp_step .01-.1]
  * 1st derivative of spectrum taken, more smoothing. [diff_window 249 for interp_step~=0.01, diff_order 2]
  * Optionally [use_threshold] peaks < [threshold_percent] * largest  peak (at [main_peak]) are rejected.
  * Optionally ([use_peaks] & something in Peaks section) peaks are matched within [peak_window] to a list.
  * Optionally peaks below an [absolute_threshold] are discarded, ignoring any baseline.
  * The next spectrum is loaded, process continues.

Troubleshooting:
================
  * "List index out of range" error may mean that the crop range doesn't overlap spectrum - check x units!''')

#Approx equal function from http://code.activestate.com/recipes/577124-approximately-equal/
def _float_approx_equal(x, y, tol=1e-18, rel=1e-7):
  if tol is rel is None:
    raise TypeError('can\'t specify both absolute and relative errors None')
  tests = []
  if tol is not None: tests.append(tol)
  if rel is not None: tests.append(rel*abs(x))
  assert tests
  return abs(x - y) <= max(tests)
def approx_in_list (a,b,tol=1e-18,rel=1e-7):
#for floats. Return 1st hit if a is within tol, rel (see above) of  item in b.  -1 if not found, like find().
  pos=-1
  for c,d in enumerate(b):
    if _float_approx_equal(a,float(d),tol,rel):
      pos=c
      break
  return pos

def Setup():
```

⋮ The setup routine writes a default configuration file.

```
def ProcessSpectrum (xdata,ydata,SpectrumNumber=""):
  global SummaryOutput
  global use_background
  global background
  vprint ("Cropping spectrum "+SpectrumNumber+" -- keeping data between " +str(ROI_min)+" and "+str(ROI_max))
  vprint ("0:"+str(xdata[0])+"; 1:"+str(xdata[1]))
  if xdata[1]<xdata[0]:#interpolation routine requires acsending order, WiRE usually does it the other way.
    vprint ("Reversing")
    xdata.reverse()
    ydata.reverse()
    vprint (xdata)
  match_size=([(max(xdata)-min(xdata))/(len(xdata)*2),abs(xdata[0]-xdata[1]),abs(xdata[-1]-xdata[-2])])
  vprint("matching using max of "+str(match_size))
  ROI_min_pos=approx_in_list(ROI_min,xdata,max(match_size))#find max, min in list.  If not, use whole list.
  ROI_max_pos=approx_in_list(ROI_max,xdata,max(match_size))
  vprint ("ROI_min_pos: "+str(ROI_min_pos)+" ROI_max_pos: "+str(ROI_max_pos))
  if ROI_min_pos<0: ROI_min_pos=0
  if ROI_max_pos<=0: ROI_max_pos=len(xdata)
  vprint ("ROI_min_pos: "+str(ROI_min_pos)+"= "+str(xdata[ROI_min_pos])+"; ROI_max_pos: "+str(ROI_max_pos)+"=
 "+str(xdata[ROI_max_pos]))
  xdata=xdata[ROI_min_pos:ROI_max_pos]
  ydata=ydata[ROI_min_pos:ROI_max_pos]
  vprint ("about to smooth")
  smoothed=[]
  smoothed=smooth(ydata,coeff)
  if File_output:
    f=open(InputFile.rpartition(".")[0]+"_"+str(SpectrumNumber)+"-smoothed."+InputFile.rpartition(".")[2],'w')
    output=zip(xdata,smoothed)
    f.write('\n'.join('%s,%s' % outitem for outitem in output))
    f.close
  if Interp_step>0:#set interp_step to 0 to avoid interpolating.
    datafunc=interp1d(xdata,smoothed,kind='cubic')
    xnew=arange(min(xdata)+Interp_step,max(xdata)-Interp_step,Interp_step)
    xnew=list(xnew)
    ynew=datafunc(xnew)
    ynew=list(ynew)
  else:
```

```
    xnew=list(xdata)
    ynew=list(smoothed)
  if use_background: ynew = [a - b for a, b in zip(ynew, background)]
  yderiv=smooth(ynew,diffcoeff)
  yderiv=list(yderiv)
  signchange=[] #detect when derivative crosses from positive to negative - i.e. peak locations.
  for count, point in enumerate(yderiv):
    if count==0:
      change=False
    elif point < 0 and yderiv[count-1] > 0:
      change = True
    else:
      change = False
    signchange.append(change)
  if File_output:
    f=open(InputFile.rpartition(".")[0]+"_"+str (SpectrumNumber)+"-deriv."+InputFile.rpartition(".")[2],'w')
    output=zip(xnew,yderiv)
    f.write('\n'.join('%s,%s' % outitem for outitem in output))
    f.close
  if graph:
    rcParams['svg.fonttype'] = 'none'
    rcParams['savefig.dpi'] = 300
    clf()
    subplot(311)
    plot(xdata,ydata,xdata,smoothed,xnew,ynew)
    if use_background:
      title('input data, smoothed input data, interpolated & bg subtracted')
    else:
      title('input data, smoothed input data, interpolated')
    subplot(312)
    plot(xnew,yderiv)
    title('derivative')
    subplot(313)
    plot(xnew,signchange)
    title('0-crossings')
    savefig(InputFile.rpartition(".")[0]+"_"+str (SpectrumNumber)+"-Graphs."+graph_type)
  Peaks=[]  #extract a list of peaks from the signchange list
  if use_threshold: #if using threshold, must find main peak and its size first
    MainPeakPos= main_peak-Peak_window
    while MainPeakPos <= main_peak+Peak_window:
      if Interp_step>0:
        MainPeakIndex=approx_in_list(MainPeakPos,xnew,Interp_step)
      else:
        MainPeakIndex=approx_in_list(MainPeakPos,xnew,max(match_size))
      if signchange[MainPeakIndex]==True:
        #we have the position of the main peak - now find the nearest turning points in the derivative
        DerivIndexLower=MainPeakIndex-1
        while (yderiv[DerivIndexLower] >yderiv[DerivIndexLower+1])and (DerivIndexLower>=ROI_min_pos):
          DerivIndexLower-=1
          if DerivIndexLower==0:break
        DerivHeightLower=yderiv[DerivIndexLower]
        DerivIndexUpper=MainPeakIndex+1
        while (yderiv[DerivIndexUpper] <yderiv[DerivIndexUpper-1])and (DerivIndexUpper<=ROI_max_pos):
          DerivIndexUpper+=1
          if DerivIndexUpper==len(yderiv)-1:break
        DerivHeightUpper=yderiv[DerivIndexUpper]
        MainPeakHeight=DerivHeightLower-DerivHeightUpper
        break
      MainPeakPos+=Interp_step
    #now found the main peak height (assuming no errors)
    if File_output:
      f=open(InputFile.rpartition(".")[0]+"_"+str(SpectrumNumber)+"-crossings."+InputFile.rpartition(".")[2],
'w')
      output=zip(xnew,signchange)
      f.write('\n'.join('%s,%s' % outitem for outitem in output))
      f.close
    vprint ("\rThreshold: Main peak found at "+str(xnew[MainPeakIndex])+" (nominal "+str(main_peak)+"), with
height "+str(MainPeakHeight)+": threshold set to "+str(threshold_percent*MainPeakHeight/100))

  for count, point in enumerate(signchange):
```

```
    if point==1:
      #Thresholding, if required
      if not (use_threshold and abs_threshold!=0):
        Peaks.append(xnew[count])
      else:
        DerivIndexLower=count-1
        while (yderiv[DerivIndexLower] >yderiv[DerivIndexLower+1])and (DerivIndexLower>=ROI_min_pos):
          DerivIndexLower-=1
        DerivHeightLower=yderiv[DerivIndexLower]
        DerivIndexUpper=count+1
        while (yderiv[DerivIndexUpper] <yderiv[DerivIndexUpper-1])and (DerivIndexUpper<=ROI_max_pos):
          DerivIndexUpper+=1
        DerivHeightUpper=yderiv[DerivIndexUpper]
        PeakHeight=DerivHeightLower-DerivHeightUpper
        if abs_threshold!=0:
          if smoothed[count]>abs_threshold>=0 or  smoothed[count]<abs_threshold<=0:
            if PeakHeight>MainPeakHeight*threshold_percent/100:
              Peaks.append(xnew[count])
              vprint ("Peak found: " +str(xnew[count])+". Derivative height "+ str(PeakHeight)+" -- Passed")
            else:
              vprint ("Peak found: " +str(xnew[count])+". Derivative height "+ str(PeakHeight)+" -- Rejected
by derivative threshold")
          else:
            vprint ("Peak found: " +str(xnew[count])+". PEak Height "+ str(smoothed[count])+" -- Rejected by
absolute threshold")
        else:
          if PeakHeight>MainPeakHeight*threshold_percent/100:
            Peaks.append(xnew[count])
            vprint ("Peak found: " +str(xnew[count])+". Derivative height "+ str(PeakHeight)+" -- Passed")
          else:
            vprint ("Peak found: " +str(xnew[count])+". Derivative height "+ str(PeakHeight)+" -- Rejected by
derivative threshold")
  if File_output:
    f=open(InputFile.rpartition(".")[0]+"_"+str(SpectrumNumber)+"-AllPeaks."+InputFile.rpartition(".")[2],'w')
    f.write('\n'.join('%s' % outitem for outitem in Peaks))
    f.close

  if Use_peaks: #trim peaks to those in PeaksList, bearing in mind tolerance
    FilterPeaks=[]
    for count, point in enumerate(Peaks):
      if approx_in_list(Peaks[count],zip(*PeaksList)[0],Peak_window,None)>-1:
        if output_height:
          NewPeak= Peaks[count],zip(*PeaksList)[1][approx_in_list(Peaks[count],zip(*PeaksList)[0],Peak_window,
None)],ynew[approx_in_list(Peaks[count],xnew,Interp_step,None)]
        else:
          NewPeak= Peaks[count],zip(*PeaksList)[1][approx_in_list(Peaks[count],zip(*PeaksList)[0],Peak_window,
None)]
        FilterPeaks.append(NewPeak)
  else:
    FilterPeaks=[]
    for count, point in enumerate(Peaks):
      if output_height:
        NewPeak= Peaks[count],zip(*PeaksList)[1][approx_in_list(Peaks[count],zip(*PeaksList)[0],Peak_window,
None)],ynew[approx_in_list(Peaks[count],xnew,Interp_step,None)]
      else:
        NewPeak= Peaks[count],zip(*PeaksList)[1][approx_in_list(Peaks[count],zip(*PeaksList)[0],Peak_window,
None)]
      FilterPeaks.append(NewPeak)
  if File_output:
    f=open(InputFile.rpartition(".")[0]+"_"+str (SpectrumNumber)+"-PEAKS."+InputFile.rpartition(".")[2],'w')
    if output_height:
      f.write('\n'.join('%s,%s,%s' % outitem for outitem in FilterPeaks))
    else:
      f.write('\n'.join('%s,%s' % outitem for outitem in FilterPeaks))
    f.close
  if Summary_file:
    if output_height:
      for FPi,FPitem in enumerate(FilterPeaks[:]):
        FPitem1,FPitem2,FPitem3=FPitem
        FilterPeaks[FPi]=FPitem3,FPitem1,FPitem2
```

```
    ZippedPeaks=zip (*FilterPeaks)
    ZippedPeaks.reverse()
    for count,data in enumerate(ZippedPeaks):
      SummaryTempList=[]
      SummaryTempList.append(SpectrumNumber)
      for item in data:
        SummaryTempList.append(item)
      SummaryOutput.append(SummaryTempList)
  if output_height:
    vprint ("Filtered Peaks Found:\n"+('\n'.join('%s,%s,%s' % outitem for outitem in FilterPeaks)))
  else:
    vprint ("Filtered Peaks Found:\n"+('\n'.join('%s,%s' % outitem for outitem in FilterPeaks)))

def ProcessFile (InputFile):
  global SummaryOutput
  try:
    f = open(InputFile, "r")
    str_in = f.read()
    f.close()
  except IOError:
    print "File %s does not exist!" % InputFile
    exit()
  str_in.replace(',','\t') #convert csv to tab-separated
  inputdata=str_in.splitlines()
  vprint (InputFile)
  vprint (" "+str(inputdata[1].count('\t'))+ " tabs (or commas) found in line 1")
  vprint (inputdata[1])
  if inputdata[1].count('\t') == 1: #we have a single spectrum (format: wavenumber \t value)
    Dimensions=1
  elif inputdata[1].count('\t')==2: # linescan? or time series - (format: xpos \t wavenumber \t value)
    Dimensions=2
  elif inputdata[1].count('\t')==3: # linescan? or area map - (format: xpos \t ypos \t wavenumber \t value)
    Dimensions=3
  else:# can't interpret data
    print "Data in file "+InputFile+" doesn't look like a spectrum, it needs to be in the format \"x<tab>y\"
 or \"x,y\" on each line with no header data, found " + str(inputdata[1].count('\t')) + "separators."
    exit()
  inputdata[:]=[datapoint.split('\t') for datapoint in inputdata] #make each data point into a 2-4 item list
  inputdata[:]=[[float(dataitem) for dataitem in datapoint] for datapoint in inputdata] #convert to floats
  if Dimensions==1:#standard 1 spectrum in a file case.
    xdata=[]
    ydata=[]
    SummaryOutput=[]
    for datapoint in inputdata:
      if ((datapoint[0]>ROI_min) and (datapoint[0]<ROI_max)): #Crops data and assigns it to x&y data lists
        xdata.append(datapoint[0])
        ydata.append(datapoint[1])
    SpectNum="-"
    ProcessSpectrum (xdata,ydata,SpectNum)
  elif Dimensions>1:#should be able to handle 2,3d files in one case if xy position is a single item
    vprint ("Multidimensional file" )
    SpectNum=""
    xdata=[]
    ydata=[]
    SummaryOutput=[]
#Last item in each datapoint is y value, penultimate: x shift, anything part of position: concatenate
    for datapoint in inputdata:
      OldSpectNum=SpectNum
      if len(datapoint)==3:
        SpectNum=str(datapoint[0])
      else:
        SpectNum=str(datapoint[0])+"_"+str(datapoint[1])
      if ((SpectNum!=OldSpectNum) and (OldSpectNum!="")):
        print OldSpectNum
        ProcessSpectrum(xdata,ydata,OldSpectNum)
        xdata[:]=[]
        ydata[:]=[]
      else:
        if ((datapoint[len(datapoint)-2]>ROI_min) and (datapoint[len(datapoint)-2]<ROI_max)):#Crops data and
 assigns it to x&y data lists
```

```
            xdata.append(datapoint[len(datapoint)-2])
            ydata.append(datapoint[len(datapoint)-1])
        print SpectNum#Change in SpectNum won't pick up the last spectrum
        ProcessSpectrum(xdata,ydata,SpectNum)
        xdata[:]=[]
        ydata[:]=[]
    if Summary_file:
        vprint (SummaryOutput)
        SummaryOutput=str(SummaryOutput)
        rx=re.compile('\], \[')
        SummaryOutput=rx.sub('\n',SummaryOutput).strip()
        rx=re.compile(', ')
        SummaryOutput=rx.sub('\t',SummaryOutput).strip()
        rx=re.compile('\'')
        SummaryOutput=rx.sub('',SummaryOutput).strip()
        rx=re.compile('\[\[')
        SummaryOutput=rx.sub('',SummaryOutput).strip()
        rx=re.compile('\]\]')
        SummaryOutput=rx.sub('',SummaryOutput).strip()
        f=open(InputFile.rpartition(".")[0]+"-Peak_summary.txt",'w')
        f.write(SummaryOutput)
        f.close

messagefile=open("peakpick-messages.txt",'w')
messagefile.write("Logging started at "+time.asctime( time.localtime(time.time()))+"\n")
```

⋮ User options are read from the command line and configuration file.

```
if Use_peaks:
    vprint ("Using Peaks:")
    vprint (PeaksList)
    vprint ("With window " +str(Peak_window))
coeff=calc_coeff ((SG_window-1)/2,SG_order) # generate smoothing coefficients
diffcoeff=calc_coeff((Diff_window-1)/2,Diff_order,1)#generate coefficients for taking 1st derivative
if use_background:
    background_file=os.path.join(os.getcwd(),background_file)
    try:
        bf=open(background_file,'r')
    except IOError:
        print "Background File %s cannot be found" %background_file
        print "looking in folder %s"
        exit()
    str_in = bf.read()
    bf.close()
    str_in.replace(',','\t') #convert csv to tab-separated
    inputdata=str_in.splitlines()
    print (background_file)
    print (" "+str(inputdata[1].count('\t'))+ " tabs (or commas) found in line 1")
    if inputdata[1].count('\t') == 1: #we have a single spectrum (format: wavenumber \t value)
        Dimensions=1
    elif inputdata[1].count('\t')==2: # linescan? or time series - (format: xpos \t wavenumber \t value)
        Dimensions=2
    elif inputdata[1].count('\t')==3: # linescan? or area map - (format: xpos \t ypos \t wavenumber \t value)
        Dimensions=3
    else:# can't interpret data
        print "Data in file "+InputFile+" doesn't look like a spectrum, it needs to be in the format \"x<tab>y\"
 or \"x,y\" on each line with no header data"
        exit()
    if Dimensions>1:
        print"Background file appears to be a map or time series - please use a single spectrum"
        exit()
    xdata=[]
    ydata=[]
    SummaryOutput=[]
    inputdata[:]=[datapoint.split('\t') for datapoint in inputdata] #make each data point into a 2-4 item list
    inputdata[:]=[[float(dataitem) for dataitem in datapoint] for datapoint in inputdata] #convert to floats
    for datapoint in inputdata:
        if ((datapoint[0]>ROI_min) and (datapoint[0]<ROI_max)):#Crops data, assigns it to x and y data lists
            xdata.append(datapoint[0])
            ydata.append(datapoint[1])
    print ("Cropping background spectrum -- keeping data between " +str(ROI_min)+" and "+str(ROI_max))
```

```
  if xdata[1]<xdata[0]:#interpolation routine requires ascending order, WiRE usually does it the other way.
    vprint ("Reversing")
    xdata.reverse()
    ydata.reverse()
    vprint (xdata)
  match_size=([(max(xdata)-min(xdata))/(len(xdata)*2),abs(xdata[0]-xdata[1]),abs(xdata[-1]-xdata[-2])])
  ROI_min_pos=approx_in_list(ROI_min,xdata,max(match_size))#try to find max & min in xlist, or use whole list
  ROI_max_pos=approx_in_list(ROI_max,xdata,max(match_size))
  vprint ("ROI_min_pos: "+str(ROI_min_pos)+" ROI_max_pos: "+str(ROI_max_pos))
  if ROI_min_pos<0: ROI_min_pos=0
  if ROI_max_pos<=0: ROI_max_pos=len(xdata)
  vprint ("ROI_min_pos: "+str(ROI_min_pos)+" ROI_max_pos: "+str(ROI_max_pos))
  xdata=xdata[ROI_min_pos:ROI_max_pos]
  ydata=ydata[ROI_min_pos:ROI_max_pos]
  vprint (xdata)
  vprint(ydata)
  datafunc=interp1d(xdata,ydata,kind='cubic')
  xnew=arange(min(xdata)+Interp_step,max(xdata)-Interp_step,Interp_step)
  xnew=list(xnew)
  background=datafunc(xnew)
  background=list(background)
  if File_output:
    f=open(background_file.rpartition(".")[0]+"-interpolated."+background_file.rpartition(".")[2],'w')
    output=zip(xnew,background)
    f.write('\n'.join('%s,%s' % outitem for outitem in output))
    f.close #end of background file processing

if InputFileName.find("*")>=0 or InputFileName.find("?")>=0:#called with wildcards
  vprint ("Filename spec passed was "+InputFileName+" -- processing multiple files")
  messagefile.write ("Filename spec passed was "+InputFileName+" -- processing multiple files")
  InputFileName=os.path.join(os.getcwd(),InputFileName)
  for InputFile in glob.glob(InputFileName):
    messagefile.write("Loading File "+InputFile)
    print "Loading File "+InputFile
    ProcessFile(InputFile)
else:
  InputFile=InputFileName
  messagefile.write("Loading File "+InputFile)
  ProcessFile(InputFile)
messagefile.write("Logging finished at "+time.asctime( time.localtime(time.time())))
messagefile.close
```

A.2 Thermal simulation spatial averaging code

This Python code was written to analyse the Thermal Analysis System device simulation output in Chapter 5, and was also applied to the thermography data in Chapter 4.

```
# TAS_neu_parse_map.py
# Chris Hodges, CDTR group 2013
# This code takes a TAS Neutral file (in TAS, File, Export, TAS Neutral File)
# which would normally have the extension .neu, crops it to just nodal position and temperatures,
# and runs a map over it, comparable to a physical map.
# read the text in showhelp below for more detail.
# Gaussian weighted averaging and nodal volume are considered.
import sys, glob, re, math, os, string
import matplotlib.pyplot as plt
import matplotlib.colors as colors
import numpy as np
from scipy.interpolate import griddata
from scipy.spatial import Voronoi,Delaunay

def ShowHelp():
  print ('''TAS_neu_parse.py
Written by Chris Hodges, CDTR group, 2013
This code takes a TAS Neutral file (in TAS, File, Export, TAS Neutral File) which would normally have the
 extension .neu
And crops it to just nodal position and temperatures (also making it comma-separated rather than space).
```

```
usage: TAS_neu_parse.py myfile.neu wildcards are OK
output will be to myfile.neu.csv
other command line arguments:
xmin, xmax, ymin, ymax, zmin, zmax, Nmin, Nmax: min and max for each coordinate and node number to plot
graph, --g: plot the data in xy - should probably specify zmin, zmax to give a single slice, this is intended
 for checking rather than publication at the moment. original data is shown in white.
xmina, xmaxa, ymina, ymaxa, zmina, zmaxa, Nmina, Nmaxa: min and max for each coordinate and node number to
 average.  If graphing, these are outlined in white
average, --a: gives the average temperature of the nodes bounded by the ?mina and ?maxa above.
omitting a min/max sets it to -/+ infinity.

MAPPING:
map, --m: Runs a map for comparison with Raman data (assumed XZ in TAS coordinates and here).  Specify this
 first of mapping arguments.  Note that mapping is done after the trim given above - so either set xmin etc.
 to larger than the mapping ROI (recommended, files are large and unsorted) or leave blank to give +/- inf.
xminm, xmaxm, yminm, ymaxm, zminm, zmaxm: min & max for each dimension in the map (set equal for a linescan)
xstep, ystepm, zstep: map step size
xsizem, zsizem: area to average over ~= spot diameter at 1/e^2 points
ysizem: depth to average over (1/e^2) but his will also be truncated by layer boundaries given by ym ~(see
 below)
ym: name, ymin, ymax triplet, can specify multiple times, e.g. ym=GaN,.2,.355 ym=Al.18GaN,-.99,.2 etc.
"map"/"--m" option must be specified before giving y
''')

def tetravol(a,b,c,d):
  '''Calculates the volume of a tetrahedron, given vertices a,b,c and d (triplets)'''
  tetravol=abs(np.dot((a-d),np.cross((b-d),(c-d))))/6
  return tetravol
def vol(vor,p):
  '''Calculate volume of 3d Voronoi cell based on point p. Voronoi diagram is passed in v.'''
  dpoints=[]
  vol=0
  for v in vor.regions[vor.point_region[p]]:
    dpoints.append(list(vor.vertices[v]))
  tri=Delaunay(np.array(dpoints),False,False,"QbB Qc Qz")
  for simplex in tri.simplices:
    vol+=tetravol(np.array(dpoints[simplex[0]]),np.array(dpoints[simplex[1]]),np.array(dpoints[simplex[2]]),np
 .array(dpoints[simplex[3]]))
  return vol
# Gauss3d calculates the  trivariate Gaussian for 1/e^2 widths xwidth, ywidth, zwidth
# at position x,y,z relative to the centre of the distibution.  Use with FWHM_axial_no_ref and Everall6
def Gauss3d (xwidth,ywidth,zwidth,x,y,z): #*width means the 1/e^2 width
  global prefactor
  sigma_x=xwidth/4 #1/e^2 == 4sigma width by definition
  sigma_y=ywidth/4
  sigma_z=zwidth/4
  TdG=prefactor*math.exp(-1*((x/sigma_x)**2+(y/sigma_y)**2+(z/sigma_z)**2))
  return TdG
def FWHM_axial_no_ref (NAa,n,PHa,lama):# ignoring refraction
  return math.sqrt(\
  ((0.88*lama)/(n-math.sqrt(n**2-NAa**2)))**2+\
  ((math.sqrt(2)*PHa)/NAa)**2)
def EverallCOG(NA,n,delta):
  global zmin,zmax
  if delta==0:
    return 0
  else:
    top=0.
    bottom=0.
    step=delta/20.
    for y in np.arange(zmin-2*step,zmax+2*step,step,dtype="float"):#here we can just set a 2-step margin.
      mm=m(y,NA,n,delta)
      II=I_m(mm,y,NA,n,delta)
      if math.isnan(mm):mm=0
      top+=mm*II*y
      bottom+=mm*II
    if top==0:
      return 0
    else:
```

```
      return(top/bottom)
def I_m(m,z,NA,n,delta):
  global zmin,zmax
  if zmin<=z<=zmax:
    I_m=math.exp(-2*m**2)
  else:#out of range or NaN
    I_m=0.
  return I_m
def m(z,NA,n,delta):
  try:
    m= math.sqrt((z**2-(n**2)*(delta**2))/((delta**2)*((NA**2)*(n**2-1)/(1-NA**2))))
    return m
  except ValueError:
    return float("nan")
def FWHM_axial (NAa,n1a,n2a,deltaa,PHa,lama):# variable list includes terms needed for refraction-based calcs.
  if deltaa==0:
    return math.sqrt(\
    ((0.88*lama)/(n1a-math.sqrt(n1a**2-NAa**2)))**2+\
    ((math.sqrt(2)*PHa)/NAa)**2)
  else:
    #first 2 lines are from the Zeiss book, 3rd is Everall's eq.7
    return math.sqrt(\
    ((0.88*lama)/(n2a-math.sqrt(n2a**2-NAa**2)))**2+\
    ((math.sqrt(2)*PHa)/NAa)**2+\
    (deltaa*(math.sqrt((NAa**2)*(n2a**2-1)/(1-NAa**2)-n2a**2)-n2a))**2)
def FWHM_lateral (NAl,nl,laml):# allowing for refraction at interface
  return 0.51*laml/(NAl/nl)
def ProcessFile (InputFile):
```

Global variables are declared.

```
  global prefactor
  FWHMe2=math.sqrt(2/math.log(2))# constant by which FWHM is multiplied to get 1/e^2 width
  # should only overwrite data from a previous run over the same simulation output
  OutputFile=InputFile+".csv"
  # not much error checking, but it's worth checking for a missing input file
  try:
    f = open(InputFile, "r")
    str_in = f.read()
    f.close()
  except IOError:
    print "File %s does not exist!" % InputFile
    exit()
  # Now simplify the file: .neu files are rather large (10s-100s of MB), so discard as much as possible
  # (there's a lot of irelevant blank/duplicate data). We want to keep lines between [NODE] and the next [END]
  # which contain 6 spaces (every other line in the NODE section) lines in between can be discarded by a regex
  print 'input file is '+str(len(str_in))+' characters long'
  inputlist=str_in.splitlines()
  print 'this is '+str(len(inputlist))+' lines'
  inputline=0
  Nodes=False#this will hold whether or not we are in the [NODES] section of the file
  f=open(OutputFile,"w") #OutputFile holds the nodal data - all of it within the selected cropped region
  f.write('Node,0,0,T,x,y,z,layer\r')#headers
  linecount=0#counts the number of output lines, just for status updates
  # the *list variables will hold the useful information about the kept nodes, as follows
  xlist=[]# x position
  ylist=[]# y position
  zlist=[]# z position
  Tlist=[]# Temperature
  Nlist=[]# TAS Node number (we can't make any assumptions about this, though they should be in order)
  if map: #set up the layer names and dims if actually mapping.  It is assumed that layers are in y.
    Nameset=zip(ynamel,yminl,ymaxl)# these lists are populated when reading in command line
    print "Nameset",Nameset#check the layers are as intended
  print "xmin, xmax, ymin, ymax, zmin, zmax, Nmin, Nmax"
  print xmin, xmax, ymin, ymax, zmin, zmax, Nmin, Nmax #Parameters within which we want to keep nodes
  while inputline<len(inputlist):# walk the input, line by line, copying out the interesting data
    if inputlist[inputline].find("NODE")>=0:# the beginning of the interesting bit
      Nodes=True
      print "found start of Nodes section"
      inputline+=1
    elif inputlist[inputline].find("END")>=0 and Nodes:# the end of the interesting bit
```

```
      print "found end of Nodes section"
      break# so no need to walk any further though the input file
    elif Nodes:#If we are in the interesting bit
      match=re.match("^([0-9]+ ){7,}$",inputlist[inputline])# >=7 blocks of digits with spaces between = line
too long to be of interest (in practice most of these numbers will be 1 or 0).
      if not(match):#so these are the interesting lines.
        op=inputlist[inputline].rstrip().replace(" ",",")
        inputitem=op.split(",")
        x=float(inputitem[4])#order of information on the line had to be determined empirically
        y=float(inputitem[5])
        z=float(inputitem[6])
        T=float(inputitem[3])
        N=float(inputitem[0])
        if (xmin<=x<=xmax)and(ymin<=y<=ymax) and (zmin<=z<=zmax)and(Nmin<=N<=Nmax):
        # test whether we actually want to keep this node
          xlist.append(x)#and store its useful info
          ylist.append(y)
          zlist.append(z)
          Tlist.append(T)
          Nlist.append(N)
        # want to check and output which layer(s) i.e. yminl, ymaxl pairs a node is in
        NameTemp=""
        for NameSetItem in Nameset:
          if NameSetItem[1]<y<NameSetItem[2]:
            #print "y",y,"nameset",Nameset
            NameTemp=NameTemp+NameSetItem[0]+" "
        NameTemp=NameTemp.strip()
        op=op+","+NameTemp+"\r"
        f.write(op)#write a line to the output file, with the relevant node information
        linecount+=1#tally the lines of output
      inputline+=1#advance the input
    else:
      inputline+=1#this is a really boring line, before the [NODES] section, so just move on
 #end while - could simply be a for loop, but originally modified the input list in place (slow)

 print 'reduced to '+str(linecount)+' lines at 1 node per line'# status update
 f.close #the file of all nodes is done with
 # plotting is optional, and not optimised, just a check really.
 #  To get a nice plot, it may be necessary to use a larger *min-*max range than otherwise desirable,
 #  so the interpolation works properly(even with linear, cubic had horrible edge-effects)
 if graph:
   xi = np.linspace(min(xlist),max(xlist),1000)# We need to interpolate the data onto a nice grid
   yi = np.linspace(min(ylist),max(ylist),1000)#  here we set up the grid.
   # set up  a nice colour map:"fire"=black-red-yellow-white
   cdict={'red':((0,0,0),(0.333,1,1),(1,1,1)), 'green':((0,0,0),(.333,0,0),(.67,1,1),(1,1,1)), 'blue'
:((0,0,0),(.67,0,0),(1,1,1))}
   my_cmap = colors.LinearSegmentedColormap('my_colormap',cdict,256)
   palette=my_cmap
   palette.set_under('b',0.0)#blue for exceptionally low values
   # grid the data.
   Ti = griddata((xlist, ylist), Tlist, (xi[None,:], yi[:,None]), method='linear')
   #plot it, note the normalisation
   CS = plt.contourf(xi,yi,Ti,50,cmap=palette,norm=colors.Normalize(vmin=max([0,np.nanmin(Ti)]),vmax=np.
nanmax(Ti),clip=True))
   plt.colorbar() # draw colorbar
   # plot data points on top
   plt.scatter(xlist,ylist,marker='+',c='b',s=3)
   #if we're doing simple averaging, plot the outline of the averaging area on top.  (Legacy)
   if average:
     plt.plot([xmina,xmina,xmaxa,xmaxa,xmina],[ymina,ymaxa,ymaxa,ymina,ymina],"w-")
   plt.show()#actually draw it
 points=zip(xlist,ylist,zlist)
 Vlist=[0.0]*len(points)# volume of Voronoi cell around point, used for weighting proportional to nodal
volume
 print "There are "+str(len(points))+" points, about to generate Voronoi diagram"
 vor=Voronoi(points)
 for i,p in enumerate(vor.points):
   out=False
   for v in vor.regions[vor.point_region[i]]:
```

```
    if v<=-1: #a point index of -1 is returned if the vertex is outside the Voronoi diagram, in this application these should be ignorable edge-cases
      out=True
      if xminm<p[0]<xmaxm and yminm<p[1]<ymaxm and zminm<p[2]<zmaxm:# we will have OOB vertices for the edge points, we don't need to care
        print "Point "+str(i)+"(TAS node "+str(Nlist[i])+") with coordinates "+str(p)+"has an out-of bounds vertex"
   if (not out):# and xminm<p[0]<xmaxm and yminm<p[1]<ymaxm and zminm<p[2]<zmaxm: #only need to calculate the volumes for points in the mapping region.
    try:
      Vlist[i]=vol(vor,i)
      #print "point "+str(i)+" with coordinates "+str(p)+" has volume "+str(Vlist[i])
    except:
      print "Error at point "+str(i)+"(TAS node "+str(Nlist[i])+") with coordinates"+str(p)+":"
      print " Voronoi diagram details"
      print " Region "+str(vor.point_region[i])+" has vertices"
      for errV in vor.regions[vor.point_region[i]]:
        print vor.vertices[errV]
      Vlist[i]=0#set the volume for this node to 0, which will mean it is ignored for averaging
pointcount=0# This is for the simple average - count the points
runningtotal=0#  keep a total
if average:
  for N in np.arange(0,len(xlist)):# walk the array(s) of points & check if they're in appropriate range
    if xmina<=xlist[N]<=xmaxa and ymina<=ylist[N]<=ymaxa and zmina<=zlist[N]<=zmaxa and Nmina<=Nlist[N]<=Nmaxa:
      pointcount+=1
      runningtotal+=Tlist[N]
  average=runningtotal/pointcount#calculate the average and output
  print "The average temperature is "+str(average)
  print "this was over the range"
  print"  "+str(xmina)+" <= x <= "+str(xmaxa)
  print"  "+str(ymina)+" <= y <= "+str(ymaxa)
  print"  "+str(zmina)+" <= z <= "+str(zmaxa)
  print"  "+str(Nmina)+" <= N <= "+str(Nmaxa)
  print str(pointcount)+" points were averaged."
if map:#walk the arrays of nodes repeatedly (x1 per point in map) straightforward,not fast or elegant
  print "About to start mapping"
  fm=open(InputFile+"_map.csv","w")# fm is the map output data file.
  fm.write("X,Y,Z")# the easy headers
  for yl in np.arange(0,len(ynamel),1):#the per-layer headers, from the name list
    fm.write(","+ynamel[yl]+" average,"+ynamel[yl]+" count")
  fm.write("\r")
  if xminm==xmaxm: #build arrays first then iterate over them - to be sure we can handle min = max cases - X
    xrange=np.empty(1)
    xrange[0]=xminm
  else:
    xrange=np.arange(xminm,xmaxm,xstepm)# the usual case, actually sweeping a range in x
  if yminm==ymaxm: # y
    yrange=np.empty(1)
    yrange[0]=yminm
  else:
    yrange=np.arange(yminm,ymaxm,ystepm)
  if zminm==zmaxm:# and z
    zrange=np.empty(1)
    zrange[0]=zminm
  else:
    zrange=np.arange(zminm,zmaxm,zstepm)
  for xm in xrange: # map over 3d
    for zm in zrange:
      for ym in yrange:
        print "xm,ym,zm",str([xm,ym,zm])
        rt= np.zeros(len(ynamel))# weighted sum of temperatures so far (per layer)
        rc= np.zeros(len(ynamel))# count of points so far (legacy, only used for status now)
        Gsum=np.zeros(len(ynamel))# sum of the weights, for calculating the average
        #delta is the sample movement, remember when plotting
        delta=-ym
        if ym>=0: #a quick hack to handle whether the spot is in air or GaN
          latsize=FWHM_lateral(NA,n1,lam)*FWHMe2
          IOR=n1
          zmin=0
```

```
            zmax=10#completely arbitrary, > top of device.
            axsize=FWHM_axial(NA,n1,n2,0,PH,lam)*FWHMe2
          else:
            latsize=FWHM_lateral(NA,n2,lam)*FWHMe2
            IOR=n2
            zmin= delta*IOR#zmin and zmax are calculated based on the definition of z_m in Everall's paper.
            zmax= delta*math.sqrt((IOR**2-NA**2)/(1-NA**2))
            axsize=FWHM_axial(NA,n1,n2,ym,PH,lam)*FWHMe2
            #axsize=FWHM_axial_no_ref(NA,n1,PH,lam)*FWHMe2
          prefactor=1/(math.sqrt((2*math.pi)**3 *(latsize/4)**2 *(axsize/4)**2 *(latsize/4)**2))
          for N,Nl in enumerate(Nlist):#work through all the nodes in the lists, by position
            if ((xm-2*latsize)<=xlist[N]<=(xm+2*latsize)) and ((zm-2*latsize)<=zlist[N]<=(zm+2*latsize)):
            # is the node in the current map spot (XZ)?
            #  treating within a radius of the spot diameter (i.e. integrate over twice the 1/e^2 width)
              for yl,yll in enumerate (ynamel):
              # in Y, things are rather different, we may have slices which are spectroscopically addressable
                if yminl[yl]<=ylist[N]<=ymaxl[yl]:#is the node in a y-slice?
                  xdiff=xm-xlist[N]# find the position within the current map spot.
                  zdiff=zm-zlist[N]
                  if delta>=0:
                    ydiff=ym-ylist[N]
                    weight=Vlist[N]*(Gauss3d (latsize,axsize,latsize,xdiff,ydiff,zdiff))
                  else:
                    ydiff=ym-ylist[N]
                    weight=Vlist[N]*(Gauss3d (latsize,axsize,latsize,xdiff,ydiff,zdiff))#calculate the 3d
 Gaussian for the spot in air
                  Gsum[yl]+=weight
                  rt[yl]+=Tlist[N]*weight
                  # This is where we multiply by the Gaussian function, taking the position of the node within
                  #  the Gaussian with respect to the current map position
                  if weight<>0: rc[yl]+=1
          fm.write(str(xm)+","+str(ym)+","+str(zm)) # output the node position (to file)
          for yl in np.arange(0,len(ynamel),1):#output the average temperature, per layer, with point count
            fm.write(","+str(rt[yl]/(Gsum[yl]))+","+str(rc[yl]))
          fm.write("\r")
        fm.write("\r")#this gives a blank line, makes finding the data in the o/p file easier
```

: User options are read from the command line, defaults are applied if options are not given.

```
  print ("Filename spec passed was "+InputFileName+" -- processing multiple files")
  InputFileName=os.path.join(os.getcwd(),InputFileName)
  for InputFile in glob.glob(InputFileName):
    print "Loading File "+InputFile
    ProcessFile(InputFile)
else: # and this is for a single file
  InputFile=InputFileName
  ProcessFile(InputFile)
```

A.3 Electroluminescence (EL) hotspot image analysis code

The Python code in this section integrates a region of interest in one or more flexible image transport system (FITS) files, for example to produce a plot of light intensity against time for a region surrounding an electroluminescence hotspot. The use of this code is described in Chapter 6, and some results analysed with it are presented in Reference 192.

```
#FITS_roi_integrate.py
#Chris Hodges, CDTR group May 2012
#This script integrates a region of interest (expressed in pixel coordinates) from one or more FITS file(s).

import ConfigParser
import sys, os, glob, re, time,imgscale,pyfits # library for working with FITS files
from math import *
from numpy import *
from pylab import *
```

```
from scipy import *
from datetime import datetime
def vprint(ToPrint):
  if verbose:
    print ToPrint
  messagefile.write (str(ToPrint)+"\n")
def ShowHelp():
  print('''FITS_roi_integrate.py
==========
Written by Chris Hodges, CDTR group 2012
Loads in one or more FITS files (assume black and white for now), looks in a region of interest specified in
 FITS_roi_integrate.cfg to find a peak, then integrates over an area also specified in the cfg file.  If the
 config file doesn't exist, run "FITS_roi_integrate.py setup" to create it.

Limited (unless "verbose") output is to STDOUT, specifying "file", "graph" or "summary", or setting
 Always_file_output, Always_summary or Always_graph in FITS_roi_integrate.cfg gives useful output to files.

Command line options:
====================
help, --h, ?: Show this help
setup:        Create the config file with default values
verbose, --v: Show lots of output
graph, --g:   Produce graphs (for debug)
file, --f:    Output to files at various stages in the process.
summary, --s: Output summary data (for map files only)
peak, --p:    Find the brightest spot and integrates around this (integration parameters to be found in .cfg
 file).  If not selected, integrates the ROI specified in the .cfg.

anything else will be interpreted as a filename, which should be last in the list of arguments
Parameters related to the region of interest in which to find a spot and the integartion to be performed once
 found can be specified in FITS_roi_integrate.cfg.
''')
def Setup():
```

: The setup routine writes a default configuration file.

```
#Processfile is the bit that does all the work - though most of the clever stuff is handled by pyfits.
def ProcessFile(inputfile):
  hdulist = pyfits.open(inputfile)
  if verbose:
    hdulist.info()
    for hdr in ['CREATOR','INSTRUME','DATE','XPIXSZ','YPIXSZ','EXPOSURE','GAIN','OFFSET','BIN_MODE','ARRAY_TY'
,'XBINNING','YBINNING','EXPTIME']: vprint(hdr+" = "+str(hdulist[0].header[hdr]))
  timestamp=hdulist[0].header['date-obs']
  vprint ("Start of exposure was at "+timestamp)
  image=hdulist[0].data
  vprint (image)
  vprint (image.shape)
  # array is 0-based, pixel coordinates usually 1-based, also y is inverted
  roi=image[image.shape[0]-ROI_max_y-1:image.shape[0]-ROI_min_y-1,ROI_min_x-1:ROI_max_x-1].copy()
  vprint (roi)
  vprint (str(image.shape[0]-ROI_max_y-1)+":"+str(image.shape[0]-ROI_min_y-1))
  if graph:#graphs are mainly for checking/debugging
    img=img_scale.linear(image)
    clf()
    imshow(img, aspect='equal')
    title(inputfile)
    savefig(inputfile.rpartition(".")[0]+".png")
    img=img_scale.linear(roi)
    clf()
    imshow(img, aspect='equal')
    title(inputfile+"-ROI: Y "+str(image.shape[0]-ROI_max_y)+":"+str(image.shape[0]-ROI_min_y)+"; X "+str(
ROI_min_x)+":"+str(ROI_max_x))
    savefig(inputfile.rpartition(".")[0]+"-ROI.png")
  if Peak:#allow for small amounts of sample movement: find brightest spot in ROI, integrate about that.
    peakpos= (unravel_index(roi.argmax(), roi.shape))#gives the peak position within ROI
    #if you get an error on the line above, check that the ROI is actually within the dimensions of the image
    #e.g. did you change the binning?
    vprint (peakpos)
    xpos=ROI_min_x+peakpos[1]
    ypos=image.shape[0]-ROI_max_y-1+peakpos[0]
```

```
    vprint (str(xpos)+","+str(ypos))
    int_region=image[ypos-integral_size_y:ypos+integral_size_y,xpos-integral_size_x:xpos+integral_size_x].copy
()
    integral=sum(int_region)
  else:
    integral=sum(roi)
  image[image.shape[0]-ROI_max_y-2:image.shape[0]-ROI_min_y,ROI_min_x-2:ROI_max_x]=32767
  image[image.shape[0]-ROI_max_y-1:image.shape[0]-ROI_min_y-1,ROI_min_x-1:ROI_max_x-1]=roi

  if graph:
    img=img_scale.linear(image)
    clf()
    imshow(img, aspect='equal')
    title(inputfile)
    savefig(inputfile.rpartition(".")[0]+".png")
    img=img_scale.linear(int_region)
    clf()
    imshow(img, aspect='equal')
    title(inputfile+"-integral_region: Y "+str(ypos-integral_size_y)+":"+str(ypos+integral_size_y)+"; X "+str(
xpos-integral_size_x)+":"+str(xpos+integral_size_x))
    savefig(inputfile.rpartition(".")[0]+"-int_region.png")
  print timestamp.replace("T"," "),integral,os.path.basename(inputfile)
  if Summary_file:#this file holds a list of intensities vs time - ideal for plotting
    summaryfile.write(timestamp.replace("T"," ")+"\t"+str(integral)+"\t"+os.path.basename(inputfile)+"\t"+str(
ypos)+"\t"+str(xpos)+"\n" )
  if graph:
    time_list.append(timestamp.replace("T"," "))
    integral_list.append(integral)
messagefile=open("peakpick-messages.txt",'w') #main
messagefile.write("Logging started at "+time.asctime( time.localtime(time.time()))+"\n")
config=ConfigParser.RawConfigParser()
File_output=False
verbose=False
graph=False
Summary_file=False
use_threshold=False
threshold_percent=0
main_peak=0
Peak=False
```

: User options are read from the command line and configuration file.

```
messagefile=open("FITS_roi_integrate-messages.txt",'w')
messagefile.write("Logging started at "+time.asctime( time.localtime(time.time()))+"\n")
outputfilename=InputFileName.replace("*","_").replace("?","_").rpartition(".")[0]
if Summary_file: summaryfile=open(outputfilename+"-summary.txt","w")
if graph:
  time_list=[]
  integral_list=[]
if InputFileName.find("*")>=0 or InputFileName.find("?")>=0:#called with wildcards
  vprint ("Filename spec passed was "+InputFileName+" -- processing multiple files")
  messagefile.write ("Filename spec passed was "+InputFileName+" -- processing multiple files")
  InputFileName=os.path.join(os.getcwd(),InputFileName)
  for InputFile in glob.glob(InputFileName):
    messagefile.write("Loading File "+InputFile)
    print "Loading File "+InputFile
    ProcessFile(InputFile)
else:
  InputFile=InputFileName
  messagefile.write("Loading File "+InputFile)
  ProcessFile(InputFile)
messagefile.close # tidying up
if Summary_file: summaryfile.close
if graph:
  clf()
  title(InputFileName)
  subplot(111)
  for i,s in enumerate(time_list): time_list[i]=datetime.strptime(time_list[i],"%Y-%m-%d %H:%M:%S")
  time_list = matplotlib.dates.date2num(time_list)
  plot_date(time_list,integral_list)
  savefig(outputfilename+"summary."+graph_type)
```

References

1. J. Bardeen and W. H. Brattain. 'The transistor, a semi-conductor triode'. *Phys. Rev.* 74 (2 July 1948), pp. 230–231. DOI: 10.1103/PhysRev.74.230.

2. H. Huff. 'From the lab to the fab: transistors to integrated circuits'. In: *Characterization and Metrology for ULSI Technology*. Ed. by D. G. Seiler, A. C. Diebold, T. J. Shaffner, R. McDonald, S. Zollner, R. P. Khosla and E. M. Secula. Vol. 683. American Institute of Physics Conference Series. Sept. 2003, pp. 3–39. DOI: 10.1063/1.1622451.

3. S. Nakamura, T. Mukai and M. Senoh. 'High-power GaN P-N junction blue-light-emitting diodes'. *Japanese Journal of Applied Physics* 30.Part 2, No. 12A (1991), pp. L1998–L2001. DOI: 10.7567/JJAP.30.L1998.

4. T. Mimura, S. Hiyamizu, T. Fujii and K. Nanbu. 'A new field-effect transistor with selectively doped GaAs/n-$Al_xGa_{1-x}As$ heterojunctions'. *Japanese Journal of Applied Physics* 19.5 (1980), pp. L225–L227. DOI: 10.7567/JJAP.19.L225.

5. M. Asif Khan, A. Bhattarai, J. N. Kuznia and D. T. Olson. 'High electron mobility transistor based on a GaN-$Al_xGa_{1-x}N$ heterojunction'. *Applied Physics Letters* 63.9 (1993), pp. 1214–1215. DOI: 10.1063/1.109775.

6. H. P. Maruska and J. J. Tietjen. 'The preparation and properties of vapor-deposited single-crystalline GaN'. *Applied Physics Letters* 15.10 (1969), pp. 327–329. DOI: 10.1063/1.1652845.

7. Y.-F. Wu, M. Moore, A. Saxler, T. Wisleder and P. Parikh. '40-W/mm double field-plated GaN HEMTs'. In: *Device Research Conference, 2006 64th*. June 2006, pp. 151–152. DOI: 10.1109/DRC.2006.305162.

8. N. Herbecq, I. Roch-Jeune, N. Rolland, D. Visalli, J. Derluyn, S. Degroote, M. Germain and F. Medjdoub. '1900 V, 1.6 $m\Omega\ cm^2$ AlN/GaN-on-Si power devices realized by local substrate removal'. *Applied Physics Express* 7.3 (2014), p. 034103. DOI: 10.7567/APEX.7.034103.

9. *Broadband RF power GaN HEMT datasheet*. Jan. 2013. URL: http://www.webcitation.org/6SGYByFvz (visited on 01/09/2014).

10. *TPH3006PS datasheet*. July 2013. URL: http://www.webcitation.org/6PuWw1l5e (visited on 09/05/2014).

11. K. V. Smith, S. Brierley, R. McAnulty, C. Tilas, D. Zarkh, M. Benedek, P. Phalon and A. Hooven. 'GaN HEMT reliability through the decade'. *ECS Transactions* 19.3 (2009), pp. 113–121. DOI: 10.1149/1.3120692.

12. P. Horowitz and W. Hill. *The Art of Electronics*. 2nd ed. Cambridge University Press, 1989. ISBN: 9780521370950.

13. J. Pankove. 'Luminescence in GaN'. *Journal of Luminescence* 7 (1973), pp. 114–126. DOI: 10.1016/0022-2313(73)90062-8.

14. U. K. Mishra, P. Parikh and Y.-F. Wu. 'AlGaN/GaN HEMTs-an overview of device operation and applications'. *Proceedings of the IEEE* 90.6 (2002), pp. 1022–1031. DOI: 10.1109/JPROC.2002.1021567.

15. *Proba-V debuts new semiconductor technology for space*. May 2013. URL: http://www.webcitation.org/6ODKPhRqh (visited on 01/01/2014).

16. M. Sugimoto, H. Ueda, M. Kanechika, N. Soejima, T. Uesugi and T. Kachi. 'Vertical device operation of AlGaN/GaN HEMTs on free-standing n-GaN substrates'. In: *Power Conversion Conference - Nagoya, 2007. PCC '07*. 2007, pp. 368–372. DOI: 10.1109/PCCON.2007.372994.

17. *Transphorm emerges from stealth to redefine energy efficiency*. Feb. 2011. URL: http://www.webcitation.org/6ODKWhwc6 (visited on 01/01/2014).

18. U. K. Mishra, L. Shen, T. Kazior and Y.-F. Wu. 'GaN-based RF power devices and amplifiers'. *Proceedings of the IEEE* 96.2 (2008), pp. 287–305. DOI: 10.1109/JPROC.2007.911060.

19. R. Willardson and E. Weber. *SiC Materials and Devices*. Semiconductors and semimetals. Elsevier Science, 1998. ISBN: 9780080864501.

20. C. Luo, D. Clarke and J. Dryden. 'The temperature dependence of the thermal conductivity of single crystal GaN films'. English. *Journal of Electronic Materials* 30.3 (2001), pp. 138–146. DOI: 10.1007/s11664-001-0007-1.

21. M. H. Wong, S. Keller, S. D. Nidhi, D. J. Denninghoff, S. Kolluri, D. F. Brown, J. Lu, N. A. Fichtenbaum, E. Ahmadi, U. Singisetti, A. Chini, S. Rajan, S. P. DenBaars, J. S. Speck and U. K. Mishra. 'N-polar GaN epitaxy and high electron mobility transistors'. *Semiconductor Science and Technology* 28.7 (2013), p. 074009. DOI: 10.1088/0268-1242/28/7/074009.

22. T. Paskova, D. Hanser and K. Evans. 'GaN substrates for III-nitride devices'. *Proceedings of the IEEE* 98.7 (2010), pp. 1324–1338. DOI: 10.1109/JPROC.2009.2030699.

23. V. Yu. Davydov, Yu. E. Kitaev, I. N. Goncharuk, A. N. Smirnov, J. Graul, O. Semchinova, D. Uffmann, M. B. Smirnov, A. P. Mirgorodsky and R. A. Evarestov. 'Phonon dispersion and Raman scattering in hexagonal GaN and AlN'. *Phys. Rev. B* 58 (Nov. 1998), pp. 12899–12907. DOI: 10.1103/PhysRevB.58.12899.

24. L. C. Lew Yan Voon, M. Willatzen, M. Cardona and N. E. Christensen. 'Terms linear in k in the band structure of wurtzite-type semiconductors'. *Phys. Rev. B* 53 (Apr. 1996), pp. 10703–10714. DOI: 10.1103/PhysRevB.53.10703.

25. S. Bloom, G. Harbeke, E. Meier and I. B. Ortenburger. 'Band structure and reflectivity of GaN'. *physica status solidi (b)* 66.1 (1974), pp. 161–168. DOI: 10.1002/pssb.2220660117.

26. M.-Z. Huang and W. Ching. 'A minimal basis semi-ab initio approach to the band structures of semiconductors'. *Journal of Physics and Chemistry of Solids* 46.8 (1985), pp. 977–995. DOI: 10.1016/0022-3697(85)90101-5.

27. W. R. L. Lambrecht, B. Segall, J. Rife, W. R. Hunter and D. K. Wickenden. 'UV reflectivity of GaN: theory and experiment'. *Phys. Rev. B* 51 (May 1995), pp. 13516–13532. DOI: 10.1103/PhysRevB.51.13516.

28. L. Vegard. 'Die Konstitution der Mischkristalle und die Raumfüllung der Atome'. German. *Zeitschrift für Physik* 5.1 (1921), pp. 17–26. DOI: 10.1007/BF01349680.

29. A. R. Denton and N. W. Ashcroft. 'Vegard's law'. *Phys. Rev. A* 43 (Mar. 1991), pp. 3161–3164. DOI: 10.1103/PhysRevA.43.3161.

30. F. Yun, M. A. Reshchikov, L. He, T. King, H. Morkoç, S. W. Novak and L. Wei. 'Energy band bowing parameter in $Al_xGa_{1-x}N$ alloys'. *Journal of Applied Physics* 92.8 (2002), pp. 4837–4839. DOI: 10.1063/1.1508420.

31. J. Wu, W. Walukiewicz, K. Yu, J. A. III, S. Li, E. Haller, H. Lu and W. J. Schaff. 'Universal bandgap bowing in group-III nitride alloys'. *Solid State Communications* 127.6 (2003), pp. 411–414. DOI: 10.1016/S0038-1098(03)00457-5.

32. K. O'Donnell, I. Fernandez-Torrente, P. Edwards and R. Martin. 'The composition dependence of the $In_xGa_{1-x}N$ bandgap'. *Journal of Crystal Growth* 269.1 (2004). Proceedings of the First ONR International Indium Nitride Workshop, pp. 100–105. DOI: 10.1016/j.jcrysgro.2004.05.040.

33. H. Taniuchi, H. Umezawa, T. Arima, M. Tachiki and H. Kawarada. 'High-frequency performance of diamond field-effect transistor'. *Electron Device Letters, IEEE* 22.8 (Aug. 2001), pp. 390–392. DOI: 10.1109/55.936353.

34. S. Rajasingam, J. Pomeroy, M. Kuball, M. Uren, T. Martin, D. C. Herbert, K. Hilton and R. Balmer. 'Micro-Raman temperature measurements for electric field assessment in active AlGaN-GaN HFETs'. *Electron Device Letters, IEEE* 25.7 (2004), pp. 456–458. DOI: 10.1109/LED.2004.830267.

35. A. Bykhovski, B. Gelmont and M. Shur. 'The influence of the strain-induced electric field on the charge distribution in GaN-AlN-GaN structure'. *Journal of Applied Physics* 74.11 (1993), pp. 6734–6739. DOI: 10.1063/1.355070.

36. F. Bernardini, V. Fiorentini and D. Vanderbilt. 'Spontaneous polarization and piezoelectric constants of III-V nitrides'. *Phys. Rev. B* 56 (Oct. 1997), R10024–R10027. DOI: 10.1103/PhysRevB.56.R10024.

37. F. Sacconi, A. Di Carlo, P. Lugli and H. Morkoc. 'Spontaneous and piezoelectric polarization effects on the output characteristics of AlGaN/GaN heterojunction modulation doped FETs'. *Electron Devices, IEEE Transactions on* 48.3 (2001), pp. 450–457. DOI: 10.1109/16.906435.

38. J. P. Ibbetson, P. T. Fini, K. D. Ness, S. P. DenBaars, J. S. Speck and U. K. Mishra. 'Polarization effects, surface states, and the source of electrons in AlGaN/GaN heterostructure field effect transistors'. *Applied Physics Letters* 77.2 (2000), pp. 250–252. DOI: 10.1063/1.126940.

39. D. Balaz, K. Kalna, M. Kuball, M. J. Uren and A. Asenov. 'Impact of the field induced polarization space-charge on the characteristics of AlGaN/GaN HEMT: self-consistent simulation study'. *physica status solidi (c)* 6.S2 (2009), S1007–S1011. DOI: 10.1002/pssc.200880858.

40. E. Kohn, I. Daumiller, M. Kunze, M. Neuburger, M. Seyboth, T. Jenkins, J. Sewell, J. Van Norstand, Y. Smorchkova and U. K. Mishra. 'Transient characteristics of GaN-based heterostructure field-effect transistors'. *Microwave Theory and Techniques, IEEE Transactions on* 51.2 (Feb. 2003), pp. 634–642. DOI: 10.1109/TMTT.2002.807687.

41. Y. Xiao, H. Shah, T. Chow and R. Gutmann. 'Analytical modeling and experimental evaluation of interconnect parasitic inductance on MOSFET switching characteristics'. In: *Applied Power Electronics Conference and Exposition, 2004. APEC '04. Nineteenth Annual IEEE*. Vol. 1. 2004, 516–521 Vol.1. DOI: 10.1109/APEC.2004.1295856.

42. Y. Cai, Y. Zhou, K. Chen and K. M. Lau. 'High-performance enhancement-mode AlGaN/GaN HEMTs using fluoride-based plasma treatment'. *Electron Device Letters, IEEE* 26.7 (July 2005), pp. 435–437. DOI: 10.1109/LED.2005.851122.

43. X. Hu, G. Simin, J. Yang, M. A. Khan, R. Gaska and M. Shur. 'Enhancement mode AlGaN/GaN HFET with selectively grown pn junction gate'. *Electronics Letters* 36.8 (Apr. 2000), pp. 753–754. DOI: 10.1049/el:20000557.

44. W. Saito, Y. Takada, M. Kuraguchi, K. Tsuda and I. Omura. 'Recessed-gate structure approach toward normally off high-voltage AlGaN/GaN HEMT for power electronics applications'. *Electron Devices, IEEE Transactions on* 53.2 (Feb. 2006), pp. 356–362. DOI: 10.1109/TED.2005.862708.

45. M. Ťapajna, S. W. Kaun, M. H. Wong, F. Gao, T. Palacios, U. K. Mishra, J. S. Speck and M. Kuball. 'Influence of threading dislocation density on early degradation in AlGaN/GaN high electron mobility transistors'. *Applied Physics Letters* 99.22 (2011), p. 223501. DOI: 10.1063/1.3663573.

46. S. W. Kaun, M. H. Wong, S. Dasgupta, S. Choi, R. Chung, U. K. Mishra and J. S. Speck. 'Effects of threading dislocation density on the gate leakage of AlGaN/GaN heterostructures for high electron mobility transistors'. *Applied Physics Express* 4.2 (2011), p. 024101. DOI: http://dx.doi.org/10.1143/APEX.4.024101.

47. *Sapphire substrates to lead future LED markets*. Aug. 2013. URL: http://www.webcitation.org/6ODKaqXrS (visited on 12/02/2014).

48. S. Pearton. *GaN and Related Materials*. GaN and Related Materials v. 1. Taylor & Francis, 1997. ISBN: 9789056995164.

49. J. Pomeroy, M. Bernardoni, A. Sarua, A. Manoi, D. Dumka, D. Fanning and M. Kuball. 'Achieving the best thermal performance for GaN-on-diamond'. In: *Compound Semiconductor Integrated Circuit Symposium (CSICS), 2013 IEEE*. 2013, pp. 1–4. DOI: 10.1109/CSICS.2013.6659210.

50. B. Moran, F. Wu, A. Romanov, U. Mishra, S. Denbaars and J. Speck. 'Structural and morphological evolution of GaN grown by metalorganic chemical vapor deposition on SiC substrates using an AlN initial layer'. *Journal of Crystal Growth* 273.1–2 (2004), pp. 38–47. DOI: 10.1016/j.jcrysgro.2004.08.012.

51. E. T. Swartz and R. O. Pohl. 'Thermal boundary resistance'. *Rev. Mod. Phys.* 61 (July 1989), pp. 605–668. DOI: 10.1103/RevModPhys.61.605.

52. H.-K. Lyeo and D. G. Cahill. 'Thermal conductance of interfaces between highly dissimilar materials'. *Phys. Rev. B* 73 (Apr. 2006), p. 144301. DOI: 10.1103/PhysRevB.73.144301.

53. E. S. Landry and A. J. H. McGaughey. 'Thermal boundary resistance predictions from molecular dynamics simulations and theoretical calculations'. *Phys. Rev. B* 80 (Oct. 2009), p. 165304. DOI: 10.1103/PhysRevB.80.165304.

54. G. H. Tang, Y. Zhao, G. X. Zhai and C. Bi. 'Phonon boundary scattering effect on thermal conductivity of thin films'. *Journal of Applied Physics* 110.4 (2011), p. 046102. DOI: 10.1063/1.3622317.

55. A. Manoi, J. Pomeroy, N. Killat and M. Kuball. 'Benchmarking of thermal boundary resistance in AlGaN/GaN HEMTs on SiC substrates: implications of the nucleation layer microstructure'. *Electron Device Letters, IEEE* 31.12 (2010). DOI: 10.1109/LED.2010.2077730.

56. A. Sarua, H. Ji, K. Hilton, D. Wallis, M. Uren, T. Martin and M. Kuball. 'Thermal boundary resistance between GaN and substrate in AlGaN/GaN electronic devices'. *Electron Devices, IEEE Transactions on* 54.12 (2007), pp. 3152–3158. DOI: 10.1109/TED.2007.908874.

57. R. Sommet, G. Mouginot, R. Quere, Z. Ouarch, S. Heckmann and M. Camiade. 'Thermal modeling and measurements of AlGaN/GaN HEMTs including thermal boundary resistance'. In: *Thermal Investigations of ICs and Systems (THERMINIC), 2010 16th International Workshop on*. 2010, pp. 1–5.

58. J. P. Freedman, J. H. Leach, E. A. Preble, Z. Sitar, R. F. Davis and J. A. Malen. 'Universal phonon mean free path spectra in crystalline semiconductors at high temperature'. *Sci. Rep.* 3 (16th Oct. 2013). DOI: 10.1038/srep02963.

59. E. A. Henriksen, S. Syed, Y. Ahmadian, M. J. Manfra, K. W. Baldwin, A. M. Sergent, R. J. Molnar and H. L. Stormer. 'Acoustic phonon scattering in a low density, high mobility AlGaN/GaN field-effect transistor'. *Applied Physics Letters* 86.25 (2005). DOI: 10.1063/1.1954893.

60. H. Harima. 'Properties of GaN and related compounds studied by means of Raman scattering'. *Journal of Physics: Condensed Matter* 14.38 (2002), R967. DOI: 10.1088/0953-8984/14/38/201.

61. T. Sander, S. Eisermann, B. K. Meyer and P. J. Klar. 'Raman tensor elements of wurtzite ZnO'. *Phys. Rev. B* 85 (Apr. 2012), p. 165208. DOI: 10.1103/PhysRevB.85.165208.

62. K. Ishioka, K. Kato, N. Ohashi, H. Haneda, M. Kitajima and H. Petek. 'The effect of n- and p-type doping on coherent phonons in GaN'. *Journal of Physics: Condensed Matter* 25.20 (2013), p. 205404. DOI: 10.1088/0953-8984/25/20/205404.

63. E. Smith and G. Dent. *Modern Raman Spectroscopy: A Practical Approach*. Wiley, 2005. ISBN: 9780471497943.

64. T. Gilson and P. Hendra. *Laser Raman Spectroscopy: a survey of interest primarily to chemists, and containing a comprehensive discussion of experiments on crystals*. Wiley-Interscience, 1970. ISBN: 9780471302001.

65. K. L. Brown and R. J. H. Clark. 'Analysis of pigmentary materials on the Vinland Map and Tartar Relation by Raman microprobe spectroscopy'. *Analytical Chemistry* 74.15 (2002), pp. 3658–3661. DOI: 10.1021/ac025610r.

66. P. Matousek and A. W. Parker. 'Bulk Raman analysis of pharmaceutical tablets'. *Applied Spectroscopy* 60.12 (2006-12-01T00:00:00), pp. 1353–1357. DOI: 10.1366/000370206779321463.

67. S. P. S. Porto, J. A. Giordmaine and T. C. Damen. 'Depolarization of Raman scattering in calcite'. *Phys. Rev.* 147 (July 1966), pp. 608–611. DOI: 10.1103/PhysRev.147.608.

68. T. C. Damen, S. P. S. Porto and B. Tell. 'Raman effect in zinc oxide'. *Phys. Rev.* 142 (Feb. 1966), pp. 570–574. DOI: 10.1103/PhysRev.142.570.

69. F. Jenkins and H. White. *Fundamentals of Optics*. 3rd. McGraw-Hill, 1957.

70. P. Yu and M. Cardona. *Fundamentals of Semiconductors: Physics and Materials Properties*. Graduate Texts in Physics. Springer, 2010. ISBN: 9783642007101.

71. G. E. Jellison, D. H. Lowndes and R. F. Wood. 'Importance of temperature-dependent optical properties for Raman-temperature measurements for silicon'. *Phys. Rev. B* 28 (Sept. 1983), pp. 3272–3276. DOI: 10.1103/PhysRevB.28.3272.

72. I. P. Herman. 'Peak temperatures from Raman Stokes/anti-Stokes ratios during laser heating by a Gaussian beam'. *Journal of Applied Physics* 109.1 (2011). DOI: 10.1063/1.3520456.

73. G. Pezzotti, H. Sueoka, A. A. Porporati, M. Manghnani and W. Zhu. 'Raman tensor elements for wurtzitic GaN and their application to assess crystallographic orientation at film/substrate interfaces'. *Journal of Applied Physics* 110.1 (2011). DOI: 10.1063/1.3606419.

74. X. Xiao-Yong, X. Sheng-Rui, Z. Jin-Cheng, L. Zhi-Yu, M. Jun-Cai, L. Zi-Yang, X. Jun-Shuai and H. Yue. 'Temperature dependences of Raman scattering in different types of GaN epilayers'. *Chinese Physics B* 21.2 (2012), p. 027803. DOI: 10.1088/1674-1056/21/2/027803.

75. A. Sarua, H. Ji, M. Kuball, M. Uren, T. Martin, K. Hilton and R. Balmer. 'Integrated micro-Raman/infrared thermography probe for monitoring of self-heating in AlGaN/GaN transistor structures'. *Electron Devices, IEEE Transactions on* 53.10 (2006), pp. 2438–2447. DOI: 10.1109/TED.2006.882274.

76. T. Batten, A. Manoi, M. J. Uren, T. Martin and M. Kuball. 'Temperature analysis of AlGaN/GaN based devices using photoluminescence spectroscopy: challenges and comparison to Raman thermography'. *Journal of Applied Physics* 107.7 (2010), p. 074502. DOI: 10.1063/1.3359651.

77. J. B. Cui, K. Amtmann, J. Ristein and L. Ley. 'Noncontact temperature measurements of diamond by Raman scattering spectroscopy'. *Journal of Applied Physics* 83.12 (1998), pp. 7929–7933. DOI: 10.1063/1.367972.

78. S. Wilhelm, B. Gröbler, M. Gluch and H. Heinz. *Confocal Laser Scanning Microscopy: Principles*. Carl Zeiss Jena. URL: http://www.webcitation.org/6PuXY8bZH.

79. N. J. Everall. ‘Modeling and measuring the effect of refraction on the depth resolution of confocal Raman microscopy’. *Applied Spectroscopy* 54.6 (Nov. 2000), pp. 773–782.

80. M. W. Davidson and M. Abramowitz. ‘Optical microscopy’. *Encyclopedia of imaging science and technology* (2002).

81. C. X. Lian, X. Y. Li and J. Liu. ‘Optical anisotropy of wurtzite GaN on sapphire characterized by spectroscopic ellipsometry’. *Semiconductor Science and Technology* 19.3 (2004), p. 417. DOI: 10.1088/0268-1242/19/3/022.

82. M. Born and E. Wolf. *Principles of Optics*. Pergamon, 1959.

83. E. Wagner, R. Dändliker and K. Spenner. *Sensors, Optical Sensors*. Sensors. Wiley, 2008. ISBN: 9783527620708.

84. M. Erdélyi and G. Gajdátsy. ‘Radial and azimuthal polarizer by means of a birefringent plate’. *Journal of Optics A: Pure and Applied Optics* 10.5 (2008), p. 055007. DOI: 10.1088/1464-4258/10/5/055007.

85. M. Erdélyi and Z. Bor. ‘Radial and azimuthal polarizers’. *Journal of Optics A: Pure and Applied Optics* 8.9 (2006), p. 737. DOI: 10.1088/1464-4258/8/9/005.

86. Y. Saito, M. Kobayashi, D. Hiraga, K. Fujita, S. Kawano, N. I. Smith, Y. Inouye and S. Kawata. ‘Z-polarization sensitive detection in micro-Raman spectroscopy by radially polarized incident light’. *Journal of Raman Spectroscopy* 39.11 (2008), pp. 1643–1648. DOI: 10.1002/jrs.1953.

87. R. Dorn, S. Quabis and G. Leuchs. ‘Sharper focus for a radially polarized light beam’. *physical review letters,* 91.23 (2003), pages. DOI: 10.1103/PhysRevLett.91.233901.

88. M. Beresna, M. Gecevičius, P. G. Kazansky and T. Gertus. ‘Radially polarized optical vortex converter created by femtosecond laser nanostructuring of glass’. *Applied Physics Letters* 98 (2011), pages. DOI: 10.1063/1.3590716.

89. Y. Kozawa and S. Sato. ‘Generation of a radially polarized laser beam by use of a conical Brewster prism’. *Opt. Lett.* 30.22 (Nov. 2005), pp. 3063–3065. DOI: 10.1364/OL.30.003063.

90. Y. Varshni. ‘Temperature dependence of the energy gap in semiconductors’. *Physica* 34.1 (1967), pp. 149–154. DOI: 10.1016/0031-8914(67)90062-6.

91. K. O’Donnell and X. Chen. ‘Temperature dependence of semiconductor band gaps’. *Applied Physics Letters* 58.25 (1991), pp. 2924–2926. DOI: 10.1063/1.104723.

92. E. Zanoni, M. Manfredi, S. Bigliardi, A. Paccagnella, P. Pisoni, C. Tedesco and C. Canali. ‘Impact ionization and light emission in AlGaAs/GaAs HEMT’s’. *Electron Devices, IEEE Transactions on* 39.8 (Aug. 1992), pp. 1849–1857. DOI: 10.1109/16.144674.

93. N. Killat, M. Montes Bajo, T. Paskova, K. R. Evans, J. Leach, X. Li, Ü. Özgür, H. Morkoç, K. D. Chabak, A. Crespo, J. K. Gillespie, R. Fitch, M. Kossler, D. E. Walker, M. Trejo, G. D. Via, J. D. Blevins and M. Kuball. ‘Reliability of AlGaN/GaN high electron mobility transistors on low dislocation density bulk

GaN substrate: implications of surface step edges'. *Applied Physics Letters* 103.19 (2013). DOI: 10.1063/1.4829062.

94. N. Killat, M. J. Uren, D. J. Wallis, T. Martin and M. Kuball. 'Origin of kink effect in AlGaN/GaN high electron mobility transistors: yellow luminescence and Fe doping'. *Applied Physics Letters* 101.15 (2012). DOI: 10.1063/1.4757993.

95. M. A. Reshchikov and H. Morkoç. 'Luminescence properties of defects in GaN'. *Journal of Applied Physics* 97.6 (Mar. 2005), p. 061301. DOI: 10.1063/1.1868059.

96. A. Anwar, S. Wu and R. T. Webster. 'Temperature dependent transport properties in GaN, $Al_xGa_{1-x}N$, and $In_xGa_{1-x}N$ semiconductors'. *Electron Devices, IEEE Transactions on* 48.3 (Mar. 2001), pp. 567–572. DOI: 10.1109/16.906452.

97. F. Gütle, V. M. Polyakov, M. Baeumler, F. Benkhelifa, S. Müller, M. Dammann, M. Cäsar, R. Quay, M. Mikulla, J. Wagner and O. Ambacher. 'Radiative inter-valley transitions as a dominant emission mechanism in AlGaN/GaN high electron mobility transistors'. *Semiconductor Science and Technology* 27.12 (2012), p. 125003. URL: http://www.csmantech.org/Digests/2010/Papers/13.2.067.pdf.

98. H. P. Zappe. 'Hot-electron electroluminescence in GaAs transistors'. *Semiconductor Science and Technology* 7.3 (1992), p. 391. DOI: 10.1088/0268-1242/7/3/020.

99. K. Turvey and J. W. Allen. 'Light emission from hot electrons in zinc selenide'. *Journal of Physics C: Solid State Physics* 6.19 (1973), p. 2887. DOI: 10.1088/0022-3719/6/19/011.

100. R. J. T. Simms, F. Gao, Y. Pei, T. Palacios, U. K. Mishra and M. Kuball. 'Electric field distribution in AlGaN/GaN high electron mobility transistors investigated by electroluminescence'. *Applied Physics Letters* 97.3 (2010). DOI: 10.1063/1.3464959.

101. P. Ivo. 'AlGaN/GaN HEMTs Reliability: Degradation Modes and Analysis'. PhD thesis. Technische Universität Berlin, 2012.

102. G. Meneghesso, G. Verzellesi, F. Danesin, F. Rampazzo, F. Zanon, A. Tazzoli, M. Meneghini and E. Zanoni. 'Reliability of GaN high-electron-mobility transistors: state of the art and perspectives'. *Device and Materials Reliability, IEEE Transactions on* 8.2 (June 2008), pp. 332–343. DOI: 10.1109/TDMR.2008.923743.

103. R. Lossy, A. Glowacki, C. Boit and J. Würfl. 'Electroluminescence characterization of AlGaN/GaN HEMTs'. *physica status solidi (c)* 6.6 (2009), pp. 1382–1385. DOI: 10.1002/pssc.200881537.

104. C. Hodges, N. Killat, S. W. Kaun, M. H. Wong, F. Gao, T. Palacios, U. K. Mishra, J. S. Speck, D. Wolverson and M. Kuball. 'Optical investigation of degradation mechanisms in AlGaN/GaN high electron mobility transistors: generation of non-radiative recombination centers'. *Applied Physics Letters* 100.11 (2012). DOI: 10.1063/1.3693427.

105. S. Sze and K. Ng. *Physics of Semiconductor Devices*. 3rd ed. Wiley, 2006. ISBN: 9788126517022.

106. D. Groom. 'Cosmic rays and other nonsense in astronomical CCD imagers'. In: *Scientific Detectors for Astronomy*. Ed. by P. Amico, J. Beletic and J. Beletic. Vol. 300. Astrophysics and Space Science Library. Springer Netherlands, 2004, pp. 81–94. ISBN: 978-1-4020-1788-9. DOI: 10.1007/1-4020-2527-0_9.

107. *Leica N Plan L 50×/0.5 datasheet*. URL: http://www.e-leica.com/Objectives.nsf/showObjectivePage?openagent&id=CCA6468BF8E142A6C1257507004FAF45 (visited on 20/05/2014).

108. *Leica N Plan EPI 50×/0.75 datasheet*. URL: http://www.e-leica.com/Objectives.nsf/showObjectivePage?openagent&id=632B586CDB8FD373C1257507004FAFEF (visited on 20/05/2014).

109. *Zeiss LD Plan-Neofluar 63×/0.75 Corr Ph2 datasheet*. URL: http://www.webcitation.org/6PuXMRJRH (visited on 20/05/2014).

110. *LMU series UV objectives*. URL: http://www.webcitation.org/6PuXlDJ4G (visited on 20/05/2014).

111. A. Chryssafis and W. Love. 'A computer-aided analysis of one-dimensional thermal transients in n-p-n power transistors'. *Solid-State Electronics* 22.3 (1979), pp. 249–256. DOI: 10.1016/0038-1101(79)90029-7.

112. W. D. Nordhaus. 'Two centuries of productivity growth in computing'. *The Journal of Economic History* 67.1 (2007), pages. DOI: 10.1017/S0022050707000058.

113. C. Hodges, J. Anaya Calvo, S. Stoffels, D. Marcon and M. Kuball. 'AlGaN/GaN field effect transistors for power electronics – Effect of finite GaN layer thickness on thermal characteristics'. *Applied Physics Letters* 103.20 (2013). DOI: 10.1063/1.4831688.

114. C. Hodges, J. Anaya Calvo and M. Kuball. *Improving depth resolution of Raman thermography on GaN HEMTs: Temperature in GaN channel and temperature gradients in the GaN buffer layer*. Poster. Bristol: UK Nitrides Consortium, Jan. 2014.

115. M. Germain, J. Derluyn, M. Van Hove, F. Medjdoub, J. Das, D. Marcon, S. Degroote, K. Cheng, M. Leys, D. Visalli, P. Srivastava, K. Geens, J. Viaene, B. Sijmus, S. Decoutere and G. Borghs. 'GaN-on-Si for power conversion'. In: *CS Mantech 2010 Digest*. May 2010, pp. 225–228. URL: http://www.csmantech.org/Digests/2010/Papers/13.2.067.pdf.

116. E. Bahat-Treidel, O. Hilt, F. Brunner, J. Würfl and G. Tränkle. 'AlGaN/GaN/AlGaN double heterojunction HEMTs on n-type SiC substrates'. *physica status solidi (c)* 7.10 (2010), pp. 2408–2411. DOI: 10.1002/pssc.200983881.

117. O. Hilt, A. Knauer, F. Brunner, E. Bahat-Treidel and J. Würfl. 'Normally-off AlGaN/GaN HFET with p-type GaN gate and AlGaN buffer'. In: *Power Semiconductor Devices IC's (ISPSD), 2010 22nd International Symposium on*. 2010, pp. 347–350.

118. E. Bahat-Treidel, O. Hilt, F. Brunner, J. Würfl and G. Tränkle. 'Punchthrough-voltage enhancement of AlGaN/GaN HEMTs using AlGaN double-heterojunction confinement'. *Electron Devices, IEEE Transactions on* 55.12 (2008), pp. 3354–3359. DOI: 10.1109/TED.2008.2006891.

119. J. Derluyn, M. Van Hove, D. Visalli, A. Lorenz, D. Marcon, P. Srivastava, K. Geens, B. Sijmus, J. Viaene, X. Kang, J. Das, F. Medjdoub, K. Cheng, S. Degroote, M. Leys, G. Borghs and M. Germain. 'Low leakage high breakdown e-mode GaN DHFET on Si by selective removal of in-situ grown Si_3N_4'. In: *Electron Devices Meeting (IEDM), 2009 IEEE International*. 2009, pp. 1–4. DOI: 10.1109/IEDM.2009.5424399.

120. P. Leturcq, J.-M. Dorkel, A. Napieralski and E. Lachiver. 'A new approach to thermal analysis of power devices'. *Electron Devices, IEEE Transactions on* 34.5 (May 1987), pp. 1147–1156. DOI: 10.1109/T-ED.1987.23057.

121. S. Vitanov. 'Simulation of High Electron Mobility Transistors'. PhD thesis. Technische Universität Wien, Fakultät für Elektrotechnik und Informationstechnik, Dec. 2010.

122. V. Palankovski, R. Schultheis and S. Selberherr. 'Simulation of power heterojunction bipolar transistors on gallium arsenide'. *Electron Devices, IEEE Transactions on* 48.6 (2001), pp. 1264–1269. DOI: 10.1109/16.925258.

123. H. Morkoç, S. Strite, G. B. Gao, M. E. Lin, B. Sverdlov and M. Burns. 'Large-band-gap SiC, III-V nitride, and II-VI ZnSe-based semiconductor device technologies'. *Journal of Applied Physics* 76.3 (1994), pp. 1363–1398. DOI: 10.1063/1.358463.

124. G. A. Slack, R. Tanzilli, R. Pohl and J. Vandersande. 'The intrinsic thermal conductivity of AlN'. *Journal of Physics and Chemistry of Solids* 48.7 (1987), pp. 641–647. DOI: 10.1016/0022-3697(87)90153-3.

125. S. W. Kaun, M. H. Wong, U. K. Mishra and J. S. Speck. 'Molecular beam epitaxy for high-performance Ga-face GaN electron devices'. *Semiconductor Science and Technology* 28.7 (2013), p. 074001. DOI: 10.1088/0268-1242/28/7/074001.

126. Y. Okada and Y. Tokumaru. 'Precise determination of lattice parameter and thermal expansion coefficient of silicon between 300 and 1500 K'. *Journal of Applied Physics* 56.2 (1984), pp. 314–320. DOI: 10.1063/1.333965.

127. J. Speight. *Lange's Handbook of Chemistry, 70th Anniversary Edition*. McGraw-Hill Professional, 2004. ISBN: 0071432205.

128. Y. Uemoto, M. Hikita, H. Ueno, H. Matsuo, H. Ishida, M. Yanagihara, T. Ueda, T. Tanaka and D. Ueda. 'Gate injection transistor (GIT) – a normally-off AlGaN/GaN power transistor using conductivity modulation'. *Electron Devices, IEEE Transactions on* 54.12 (2007), pp. 3393–3399. DOI: 10.1109/TED.2007.908601.

129. R. Chu, Z. Chen, S. P. DenBaars and U. K. Mishra. 'V-gate GaN HEMTs with engineered buffer for normally off operation'. *Electron Device Letters, IEEE* 29.11 (2008), pp. 1184–1186. DOI: 10.1109/LED.2008.2004721.

130. F. Medjdoub, J. Derluyn, K. Cheng, M. Leys, S. Degroote, D. Marcon, D. Visalli, M. Van Hove, M. Germain and G. Borghs. 'Low on-resistance high-breakdown normally off AlN/GaN/AlGaN DHFET on Si substrate'. *Electron Device Letters, IEEE* 31.2 (2010), pp. 111–113. DOI: 10.1109/LED.2009.2037719.

131. S. Stoffels and D. Marcon. Private communication by email. IMEC. Feb. 2012.

132. B. C. Daly, H. J. Maris, A. V. Nurmikko, M. Kuball and J. Han. 'Optical pump-and-probe measurement of the thermal conductivity of nitride thin films'. *Journal of Applied Physics* 92.7 (2002), pp. 3820–3824. DOI: 10.1063/1.1505995.

133. W. Liu and A. A. Balandin. 'Temperature dependence of thermal conductivity of $Al_xGa_{1-x}N$ thin films measured by the differential 3 omega technique'. *Applied Physics Letters* 85.22 (2004), pp. 5230–5232. DOI: 10.1063/1.1829168.

134. M. S. Liu, L. A. Bursill and S. Prawer. 'Optical phonons of hexagonal $Al_xGa_{1-x}N$: simulation and experiment'. *Journal of Applied Physics* 90.4 (2001), pp. 1761–1767. DOI: 10.1063/1.1384508.

135. F. Demangeot, J. Groenen, J. Frandon, M. A. Renucci, O. Briot, S. Clur and R. L. Aulombard. 'Coupling of GaN- and AlN-like longitudinal optic phonons in $Ga_{1-x}Al_xN$ solid solutions'. *Applied Physics Letters* 72.21 (1998), pp. 2674–2676. DOI: 10.1063/1.121095.

136. V. Yu. Davydov, I. N. Goncharuk, A. N. Smirnov, A. E. Nikolaev, W. V. Lundin, A. S. Usikov, A. A. Klochikhin, J. Aderhold, J. Graul, O. Semchinova and H. Harima. 'Composition dependence of optical phonon energies and Raman line broadening in hexagonal $Al_xGa_{1-x}N$ alloys'. *Phys. Rev. B* 65 (Mar. 2002), p. 125203. DOI: 10.1103/PhysRevB.65.125203.

137. B. Jusserand and J. Sapriel. 'Raman investigation of anharmonicity and disorder-induced effects in $Ga_{1-x}Al_xAs$ epitaxial layers'. *Phys. Rev. B* 24 (Dec. 1981), pp. 7194–7205. DOI: 10.1103/PhysRevB.24.7194.

138. P. Parayanthal and F. H. Pollak. 'Raman scattering in alloy semiconductors: "spatial correlation" model'. *Phys. Rev. Lett.* 52 (May 1984), pp. 1822–1825. DOI: 10.1103/PhysRevLett.52.1822.

139. K. Cheng, M. Leys, S. Degroote, B. Daele, S. Boeykens, J. Derluyn, M. Germain, G. Tendeloo, J. Engelen and G. Borghs. 'Flat GaN epitaxial layers grown on Si(111) by metalorganic vapor phase epitaxy using step-graded AlGaN intermediate layers'. *Journal of Electronic Materials* 35.4 (2006), pp. 592–598. DOI: 10.1007/s11664-006-0105-1.

140. G. Riedel, J. Pomeroy, K. Hilton, J. Maclean, D. Wallis, M. Uren, T. Martin, U. Forsberg, A. Lundskog, A. Kakanakova-Georgieva, G. Pozina, E. Janzén, R. Lossy, R. Pazirandeh, F. Brunner, J. Würfl and M. Kuball. 'Reducing thermal resistance of AlGaN/GaN electronic devices using novel nucleation layers'. *Electron Device Letters, IEEE* 30.2 (2009), pp. 103–106.

141. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing (2nd Ed.)* New York, NY, USA: Cambridge University Press, 1992. ISBN: 0-521-43108-5.

142. T. O'Haver. *A Pragmatic Introduction to Signal Processing with applications in chemical analysis*. http://www.webcitation.org/6MFXv679h. Dec. 2013.

143. A. Savitzky and M. J. E. Golay. 'Smoothing and differentiation of data by simplified least squares procedures.' *Analytical Chemistry* 36.8 (1964), pp. 1627–1639. DOI: 10.1021/ac60214a047.

144. J. Kuzmik, P. Javorka, A. Alam, M. Marso, M. Heuken and P. Kordos. 'Determination of channel temperature in AlGaN/GaN HEMTs grown on sapphire and silicon substrates using DC characterization method'. *Electron Devices, IEEE Transactions on* 49.8 (2002), pp. 1496–1498. DOI: 10.1109/TED.2002.801430.

145. R. J. T. Simms, J. Pomeroy, M. Uren, T. Martin and M. Kuball. 'Channel temperature determination in high-power AlGaN/GaN HFETs using electrical methods and Raman spectroscopy'. *Electron Devices, IEEE Transactions on* 55.2 (2008), pp. 478–482. DOI: 10.1109/TED.2007.913005.

146. D. I. Florescu, V. M. Asnin, F. H. Pollak, A. M. Jones, J. C. Ramer, M. J. Schurman and I. Ferguson. 'Thermal conductivity of fully and partially coalesced lateral epitaxial overgrown GaN/sapphire (0001) by scanning thermal microscopy'. *Applied Physics Letters* 77.10 (2000), pp. 1464–1466. DOI: 10.1063/1.1308057.

147. B. A. Danilchenko, T. Paszkiewicz, S. Wolski, A. Jeżowski and T. Plackowski. 'Heat capacity and phonon mean free path of wurtzite GaN'. *Applied Physics Letters* 89.6 (2006), p. 061901. DOI: 10.1063/1.2335373.

148. J. Zou, D. Kotchetkov, A. A. Balandin, D. I. Florescu and F. H. Pollak. 'Thermal conductivity of GaN films: effects of impurities and dislocations'. *Journal of Applied Physics* 92.5 (2002), pp. 2534–2539. DOI: 10.1063/1.1497704.

149. D. T. Morelli, J. P. Heremans and G. A. Slack. 'Estimation of the isotope effect on the lattice thermal conductivity of group IV and group III-V semiconductors'. *Phys. Rev. B* 66 (Nov. 2002), p. 195304. DOI: 10.1103/PhysRevB.66.195304.

150. E. Ziambaras and P. Hyldgaard. 'Thermal transport in SiC nanostructures'. *Materials Science and Engineering: C* 25.5–8 (2005), pp. 635–640. DOI: 10.1016/j.msec.2005.06.044.

151. E. Ziambaras and P. Hyldgaard. 'Phonon Knudsen flow in nanostructured semiconductor systems'. *Journal of Applied Physics* 99.5 (2006), p. 054303. DOI: 10.1063/1.2175474.

152. Y. Ma. 'Size-dependent thermal conductivity in nanosystems based on non-Fourier heat transfer'. *Applied Physics Letters* 101.21 (2012), p. 211905. DOI: 10.1063/1.4767337.

153. F. Meng, J. Zhang, H. Zhou, J. Ma, J. Xue, L. Dang, L. Zhang, M. Lu, S. Ai, X. Li and Y. Hao. 'Transport characteristics of AlGaN/GaN/AlGaN double heterostructures with high electron mobility'. *Journal of Applied Physics* 112.2 (2012). DOI: 10.1063/1.4739408.

154. C. Hodges, J. Pomeroy and M. Kuball. 'Probing temperature gradients within the GaN buffer layer of AlGaN/GaN high electron mobility transistors with

Raman thermography'. *Journal of Applied Physics* 115.6 (2014). DOI: 10.1063/1.4865296.

155. K. P. J. Williams, G. D. Pitt, D. N. Batchelder and B. J. Kip. 'Confocal Raman microspectroscopy using a stigmatic spectrograph and CCD detector'. *Applied Spectroscopy* 48 (1994), pp. 232–235.

156. J. Christofferson and A. Shakouri. 'Thermoreflectance based thermal microscope'. *Review of Scientific Instruments* 76.2 (2005). DOI: 10.1063/1.1850632.

157. A. Soudi, R. D. Dawson and Y. Gu. 'Quantitative heat dissipation characteristics in current-carrying GaN nanowires probed by combining scanning thermal microscopy and spatially resolved raman spectroscopy'. *ACS Nano* 5.1 (2011), pp. 255–262. DOI: 10.1021/nn102818s.

158. N. Killat, M. Kuball, T. Chou, U. Chowdhury and J. Jimenez. 'Temperature assessment of AlGaN/GaN HEMTs: a comparative study by Raman, electrical and IR thermography'. In: *Reliability Physics Symposium (IRPS), 2010 IEEE International*. 2010, pp. 528–531. DOI: 10.1109/IRPS.2010.5488777.

159. J. Joh, U. Chowdhury, T.-M. Chou, H.-Q. Tserng and J. Jimenez. 'Method for estimation of the channel temperature of GaN high electron mobility transistors'. In: *ROCS Workshop, 2007.[Reliability of Compound Semiconductors Digest]*. Oct. 2007, pp. 87–101. DOI: 10.1109/ROCS.2007.4391064.

160. B. Schwitter, A. Parker, S. Mahon, A. Fattorini and M. Heimlich. 'Impact of bias and device structure on gate junction temperature in AlGaN/GaN-on-Si HEMTs'. *Electron Devices, IEEE Transactions on* PP.99 (2014). In press, pp. 1–1. DOI: 10.1109/TED.2014.2311660.

161. W. Roesch. 'Thermo-reliability relationships of GaAs ics'. In: *Gallium Arsenide Integrated Circuit (GaAs IC) Symposium, 1988. Technical Digest 1988., 10th Annual IEEE*. Nov. 1988, pp. 61–64. DOI: 10.1109/GAAS.1988.11024.

162. J. J. Cargille. 'Immersion oil and the microscope (revised 1985)'. *New York Microscopical Society Yearbook* (1964). URL: http://www.webcitation.org/6ODTLL9u1.

163. S. M. Mansfield and G. S. Kino. 'Solid immersion microscope'. *Applied Physics Letters* 57.24 (1990), pp. 2615–2616. DOI: 10.1063/1.103828.

164. Q. Wu, G. D. Feke, R. D. Grober and L. P. Ghislain. 'Realization of numerical aperture 2.0 using a gallium phosphide solid immersion lens'. *Applied Physics Letters* 75.26 (1999), pp. 4064–4066. DOI: 10.1063/1.125537.

165. D. Fletcher, K. B. Crozier, K. Guarini, S. Minne, G. S. Kino, C. F. Quate and K. E. Goodson. 'Microfabricated silicon solid immersion lens'. *Microelectromechanical Systems, Journal of* 10.3 (Sept. 2001), pp. 450–459. DOI: 10.1109/84.946806.

166. S. B. Ippolito, B. B. Goldberg and M. S. Ünlü. 'Theoretical analysis of numerical aperture increasing lens microscopy'. *Journal of Applied Physics* 97.5 (2005). DOI: 10.1063/1.1858060.

167. P. Scholz, C. Gallrapp, U. Kerst, T. Lundquist and C. Boit. 'Optimizing focused ion beam created solid immersion lenses in bulk silicon using design of experiments'. *Microelectronics Reliability* 50.9–11 (2010). 21st European Symposium on the Reliability of Electron Devices, Failure Physics and Analysis, pp. 1441–1445. DOI: 10.1016/j.microrel.2010.07.031.

168. M. Kuball, J. W. Pomeroy, M. Montes Bajo, M. Silvestri, M. J. Uren and N. Killat. 'Junction temperature measurements and reliability of GaN FETs'. In: vol. 8625. 2013, pages. DOI: 10.1117/12.2007817.

169. G. M. Lerman and U. Levy. 'Effect of radial polarization and apodization on spot size under tight focusing conditions'. *Opt. Express* 16.7 (Mar. 2008), pp. 4567–4581. DOI: 10.1364/OE.16.004567.

170. L. Novotny and S. J. Stranick. 'Near-field optical microscopy and spectroscopy with pointed probes'. *Annual Review of Physical Chemistry* 57 (2006). DOI: 10.1146/annurev.physchem.56.092503.141236.

171. K. D. Alexander and Z. D. Schultz. 'Tip-enhanced Raman detection of antibody conjugated nanoparticles on cellular membranes'. *Analytical Chemistry* 84.17 (2012), pp. 7408–7414. DOI: 10.1021/ac301739k.

172. K. Youngworth and T. Brown. 'Focusing of high numerical aperture cylindrical-vector beams'. *Opt. Express* 7.2 (July 2000), pp. 77–87. DOI: 10.1364/OE.7.000077.

173. R. H. Jordan and D. G. Hall. 'Free-space azimuthal paraxial wave equation: the azimuthal Bessel–Gauss beam solution'. *Opt. Lett.* 19.7 (Apr. 1994), pp. 427–429. DOI: 10.1364/OL.19.000427.

174. C. Sheppard, M. Sharma and A. Arbouet. 'Axial apodizing filters for confocal imaging'. *Optik* 111.8 (2000), pp. 347–354.

175. Q. Zhan and J. Leger. 'Focus shaping using cylindrical vector beams'. *Opt. Express* 10.7 (Apr. 2002), pp. 324–331. DOI: 10.1364/OE.10.000324.

176. R. N. Smartt. 'The production of high-quality optical surfaces on calcite'. *Journal of Scientific Instruments* 38.4 (1961), p. 165. DOI: 10.1088/0950-7671/38/4/419.

177. Y. Yin-Tang, H. Ru and Wang-Ping. 'Raman analysis of defects in n-type 4H-SiC'. *Chinese Physics B* 17.9 (2008), p. 3459. DOI: 10.1088/1674-1056/17/9/053.

178. S. Nakashima, A. Wada and Z. Inoue. 'Raman scattering from anisotropic phonon modes in SiC polytypes'. *Journal of the Physical Society of Japan* 56.9 (1987), pp. 3375–3380. DOI: 10.1143/JPSJ.56.3375.

179. V. Yu. Davydov, N. S. Averkiev, I. N. Goncharuk, D. K. Nelson, I. P. Nikitina, A. S. Polkovnikov, A. N. Smirnov, M. A. Jacobson and O. K. Semchinova. 'Raman and photoluminescence studies of biaxial strain in GaN epitaxial layers grown on 6H–SiC'. *Journal of Applied Physics* 82.10 (1997), pp. 5097–5102. DOI: 10.1063/1.366310.

180. D. W. Feldman, J. H. Parker, W. J. Choyke and L. Patrick. 'Phonon dispersion curves by Raman scattering in SiC, polytypes 3C, 4H, 6H, 15R, and 21R'. *Phys. Rev.* 173 (Sept. 1968), pp. 787–793. DOI: 10.1103/PhysRev.173.787.

181. M. Tadjer, T. Anderson, K. Hobart, T. Feygelson, J. Caldwell, C. Eddy, F. Kub, J. Butler, B. Pate and J. Melngailis. 'Reduced self-heating in AlGaN/GaN HEMTs using nanocrystalline diamond heat-spreading films'. *Electron Device Letters, IEEE* 33.1 (2012), pp. 23–25. DOI: 10.1109/LED.2011.2171031.

182. H. J. Park, C. Park, S. Yeo, S. Kang, M. Mastro, O. Kryliouk and T. J. Anderson. 'Epitaxial strain energy measurements of GaN on sapphire by Raman spectroscopy'. *physica status solidi (c)* 2.7 (2005), pp. 2446–2449. DOI: 10.1002/pssc.200461513.

183. C. Röder, F. Lipski, F. Habel, G. Leibiger, M. Abendroth, C. Himcinschi and J. Kortus. 'Raman spectroscopic characterization of epitaxially grown GaN on sapphire'. *Journal of Physics D: Applied Physics* 46.28 (2013), p. 285302. DOI: 10.1088/0022-3727/46/28/285302.

184. G. Nootz, A. Schulte, L. Chernyak, A. Osinsky, J. Jasinski, M. Benamara and Z. Liliental-Weber. 'Correlations between spatially resolved Raman shifts and dislocation density in GaN films'. *Applied Physics Letters* 80.8 (Feb. 2002), pp. 1355–1357. DOI: 10.1063/1.1449523.

185. C. Mion, J. F. Muth, E. A. Preble and D. Hanser. 'Accurate dependence of gallium nitride thermal conductivity on dislocation density'. *Applied Physics Letters* 89.9, 092123 (2006), pages. DOI: http://dx.doi.org/10.1063/1.2335972.

186. J. Cho, E. Bozorg-Grayeli, D. Altman, M. Asheghi and K. E. Goodson. 'Low thermal resistances at GaN–SiC interfaces for HEMT technology'. *Electron Device Letters, IEEE* 33.3 (2012), pp. 378–380. DOI: 10.1109/LED.2011.2181481.

187. S. Kaiser, M. Jakob, J. Zweck, W. Gebhardt, O. Ambacher, R. Dimitrov, A. T. Schremer, J. A. Smart and J. R. Shealy. 'Structural properties of AlGaN/GaN heterostructures on Si(111) substrates suitable for high-electron mobility transistors'. *Journal of Vacuum Science & Technology B* 18.2 (2000), pp. 733–740. DOI: 10.1116/1.591268.

188. R. B. Simon, J. W. Pomeroy and M. Kuball. 'Diamond micro-raman thermometers for accurate gate temperature measurements'. *Applied Physics Letters* 104.21, 213503 (2014), pages. DOI: http://dx.doi.org/10.1063/1.4879849.

189. S. Chowdhury, B. L. Swenson, M. H. Wong and U. K. Mishra. 'Current status and scope of gallium nitride-based vertical transistors for high-power electronics application'. *Semiconductor Science and Technology* 28.7 (2013), p. 074014. URL: http://stacks.iop.org/0268-1242/28/i=7/a=074014.

190. H. Mitlehner, W. Bartsch, K.-O. Dohnke, P. Friedrichs, R. Kaltschmidt, U. Weinert, B. Weis and D. Stephani. 'Dynamic characteristics of high voltage 4h-sic vertical jfets'. In: *Power Semiconductor Devices and ICs, 1999. ISPSD '99. Proceedings., The 11th International Symposium on.* 1999, pp. 339–342. DOI: 10.1109/ISPSD.1999.764129.

191. J. Steeds, G. Evans, L. Danks, S. Furkert, W. Voegeli, M. Ismail and F. Carosella. 'Transmission electron microscope radiation damage of 4H and 6H SiC studied by photoluminescence spectroscopy'. *Diamond and Related Materials* 11.12 (2002), pp. 1923–1945. DOI: http://dx.doi.org/10.1016/S0925-9635(02)00212-1.

192. M. Montes Bajo, C. Hodges, M. Uren and M. Kuball. 'On the link between electroluminescence, gate current leakage, and surface defects in AlGaN/GaN high electron mobility transistors upon off-state stress'. *Applied Physics Letters* 101.3 (2012), p. 033508. DOI: 10.1063/1.4737904.

193. F. Reif. *Fundamentals of Statistical and Thermal Physics*. New York: McGraw Hill, 1965.

194. N. Shigekawa, K. Shiojima and T. Suemitsu. 'Electroluminescence characterization of AlGaN/GaN high-electron-mobility transistors'. *Applied Physics Letters* 79.8 (2001), pp. 1196–1198. DOI: 10.1063/1.1398332.

195. E. Zanoni, G. Meneghesso, G. Verzellesi, F. Danesin, M. Meneghini, F. Rampazzo, A. Tazzoli and F. Zanon. 'A review of failure modes and mechanisms of GaN-based HEMTs'. In: *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*. Dec. 2007, pp. 381–384. DOI: 10.1109/IEDM.2007.4418952.

196. E. Zanoni, F. Danesin, M. Meneghini, A. Cetronio, C. Lanzieri, M. Peroni and G. Meneghesso. 'Localized damage in AlGaN/GaN HEMTs induced by reverse-bias testing'. *Electron Device Letters, IEEE* 30.5 (May 2009), pp. 427–429. DOI: 10.1109/LED.2009.2016440.

197. J. Joh and J. A. del Alamo. 'Mechanisms for electrical degradation of GaN high-electron mobility transistors'. In: *Electron Devices Meeting, 2006. IEDM '06. International*. Dec. 2006, pp. 1–4. DOI: 10.1109/IEDM.2006.346799.

198. D. Cullen, D. Smith, A. Stocco, G. Meneghesso, E. Zanoni and M. McCartney. 'Characterization of localized degradation in reverse-biased GaN HEMTs by scanning transmission electron microscopy and electron holography'. *Microscopy and Microanalysis* 16 (July 2010), pp. 800–801. DOI: 10.1017/S1431927610056722.

199. E. Zanoni, M. Meneghini, A. Chini, D. Marcon and G. Meneghesso. 'AlGaN/GaN-based HEMTs failure physics and reliability: mechanisms affecting gate edge and Schottky junction'. *Electron Devices, IEEE Transactions on* 60.10 (Oct. 2013), pp. 3119–3131. DOI: 10.1109/TED.2013.2271954.

200. D. J. Chen, J. J. Xue, B. Liu, H. Lu, Z. L. Xie, P. Han, R. Zhang, Y. D. Zheng, Y. C. Kong and J. J. Zhou. 'Observation of hole accumulation at the interface of an undoped InGaN/GaN heterostructure'. *Applied Physics Letters* 95.1, 012112 (2009), pages. DOI: 10.1063/1.3176443.

201. K. Kunihiro, K. Kasahara, Y. Takahashi and Y. Ohno. 'Experimental evaluation of impact ionization coefficients in GaN'. *Electron Device Letters, IEEE* 20.12 (Dec. 1999), pp. 608–610. DOI: 10.1109/55.806100.

202. N. Killat, M. Ťapajna, M. Faqir, T. Palacios and M. Kuball. 'Evidence for impact ionisation in AlGaN/GaN HEMTs with InGaN back-barrier'. *Electronics Letters* 47.6 (2011), pp. 405–406. DOI: 10.1049/el.2010.7540.

203. B. Brar, K. Boutros, R. DeWarnes, V. Tilak, R. Shealy and L. Eastman. 'Impact ionization in high performance AlGaN/GaN HEMTs'. In: *High Performance Devices, 2002. Proceedings. IEEE Lester Eastman Conference on*. Aug. 2002, pp. 487–491. DOI: 10.1109/LECHPD.2002.1146791.

204. M. Ťapajna, R. Simms, Y. Pei, U. Mishra and M. Kuball. 'Integrated optical and electrical analysis: identifying location and properties of traps in AlGaN/GaN HEMTs during electrical stress'. *Electron Device Letters, IEEE* 31.7 (July 2010), pp. 662–664. DOI: 10.1109/LED.2010.2047092.

205. X. Cao, P. Sandvik, S. LeBoeuf and S. Arthur. 'Defect generation in InGaN/GaN light-emitting diodes under forward and reverse electrical stresses'. *Microelectronics Reliability* 43.12 (2003), pp. 1987–1991. DOI: 10.1016/j.microrel.2003.06.001.

206. J. Mickevicius, R. Aleksiejunas, M. Shur, J. Zhang, Q. Fareed, R. Gaska and G. Tamulaitis. 'Lifetime of nonequilibrium carriers in AlGaN epilayers with high Al molar fraction'. In: *High Performance Devices, 2004. Proceedings. IEEE Lester Eastman Conference on*. Aug. 2004, pp. 72–77. DOI: 10.1109/LECHPD.2004.1549674.

207. S. Nakamura. 'The roles of structural imperfections in InGaN-based blue light-emitting diodes and laser diodes'. *Science* 281.5379 (1998), pp. 956–961. DOI: 10.1126/science.281.5379.956.

208. L. Chernyak, A. Osinsky, H. Temkin, J. W. Yang, Q. Chen and M. Asif Khan. 'Electron beam induced current measurements of minority carrier diffusion length in gallium nitride'. *Applied Physics Letters* 69.17 (1996), pp. 2531–2533. DOI: 10.1063/1.117729.

209. L. G. Baird. 'Near field imaging of gallium nitride nanowires for characterization of minority carrier diffusion'. MS Thesis. Naval postgraduate school, Dec. 2009.

210. J. W. Yang, C. J. Sun, Q. Chen, M. Z. Anwar, M. Asif Khan, S. A. Nikishin, G. A. Seryogin, A. V. Osinsky, L. Chernyak, H. Temkin, C. Hu and S. Mahajan. 'High quality GaN–InGaN heterostructures grown on (111) silicon substrates'. *Applied Physics Letters* 69.23 (1996), pp. 3566–3568. DOI: 10.1063/1.117247.

211. M. M. Bajo, H. Sun, M. J. Uren and M. Kuball. 'Time evolution of off-state degradation of AlGaN/GaN high electron mobility transistors'. *Applied Physics Letters* 104.22, 223506 (2014), pages. DOI: 10.1063/1.4881637.

212. M. Meneghini, A. Stocco, M. Bertin, D. Marcon, A. Chini, G. Meneghesso and E. Zanoni. 'Time-dependent degradation of AlGaN/GaN high electron mobility transistors under reverse bias'. *Applied Physics Letters* 100.3 (2012). DOI: 10.1063/1.3678041.

213. R. Degraeve, G. Groeseneken, R. Bellens, J. Ogier, M. Depas, P. Roussel and H. Maes. 'New insights in the relation between electron trap generation and the statistical properties of oxide breakdown'. *Electron Devices, IEEE Transactions on* 45.4 (Apr. 1998), pp. 904–911. DOI: 10.1109/16.662800.

214. P. Ivo, A. Glowacki, R. Pazirandeh, E. Bahat-Treidel, R. Lossy, J. Wurfl, C. Boit and G. Trankle. 'Influence of GaN cap on robustness of AlGaN/GaN HEMTs'. In: *Reliability Physics Symposium, 2009 IEEE International*. Apr. 2009, pp. 71–75. DOI: 10.1109/IRPS.2009.5173226.

215. D. Marcon, T. Kauerauf, F. Medjdoub, J. Das, M. Van Hove, P. Srivastava, K. Cheng, M. Leys, R. Mertens, S. Decoutere, G. Meneghesso, E. Zanoni and G. Borghs. 'A comprehensive reliability investigation of the voltage-, temperature- and device geometry-dependence of the gate degradation on state-of-the-art GaN-on-Si HEMTs'. In: *Electron Devices Meeting (IEDM), 2010 IEEE International*. Dec. 2010, pp. 20.3.1–20.3.4. DOI: 10.1109/IEDM.2010.5703398.

216. M. Ťapajna, U. K. Mishra and M. Kuball. 'Importance of impurity diffusion for early stage degradation in AlGaN/GaN high electron mobility transistors upon electrical stress'. *Applied Physics Letters* 97.2 (2010), p. 023503. DOI: 10.1063/1.3460529.

217. P. Ivo, A. Glowacki, E. Bahat-Treidel, R. Lossy, J. Würfl, C. Boit and G. Tränkle. 'Comparative study of AlGaN/GaN HEMTs robustness versus buffer design variations by applying electroluminescence and electrical measurements'. *Microelectronics Reliability* 51.2 (2011), pp. 217–223. DOI: DOI:10.1016/j.microrel.2010.09.029.

218. A. Chini, V. Di Lecce, M. Esposto, G. Meneghesso and E. Zanoni. 'Evaluation and numerical simulations of GaN HEMTs electrical degradation'. *Electron Device Letters, IEEE* 30.10 (Oct. 2009), pp. 1021–1023. DOI: 10.1109/LED.2009.2029875.

219. J. Neugebauer and C. G. Van de Walle. 'Gallium vacancies and the yellow luminescence in GaN'. *Applied Physics Letters* 69.4 (1996), pp. 503–505. DOI: 10.1063/1.117767.

220. H. Sun, M. Montes Bajo, M. J. Uren and M. Kuball. 'Implications of gate edge electric field in AlGaN/GaN high electron mobility transistors during OFF-state degradation'. *Microelectronics Reliability* (2014). Accepted.

221. A. Schneider, D. Wolverson, K. Sebald, C. Hodges, M. Kuball and T. Voss. 'Analysis of strained surface layers of ZnO single crystals after irradiation with intense femtosecond laser pulses'. *Applied Physics Letters* 102.21 (2013), p. 211904. DOI: 10.1063/1.4807923.

222. J. Möreke, C. Hodges, L. L. Mears, M. J. Uren, R. M. Richardson and M. Kuball. 'Liquid crystal electrography: electric field mapping and detection of peak electric field strength in AlGaN/GaN high electron mobility transistors'. *Microelectronics Reliability* 54.5 (2014), pp. 921–925. DOI: 10.1016/j.microrel.2014.01.006.

223. C. Hodges, N. Killat and M. Kuball. *AlGaN/GaN HEMTs Device Degradation: Hot Carrier and Blue Defect-Related Electroluminescence*. Oral Presentation. Glasgow: 9th International Conference on Nitride Semiconductors, 2011.